Traffic Accident Scene Recognition with FMCW Radar and Vision Transformer

Runwei Guan
University of Liverpool
5 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Shanliang Yao
University of Liverpool
4 PUBLICATIONS **0** CITATIONS
SEE PROFILE

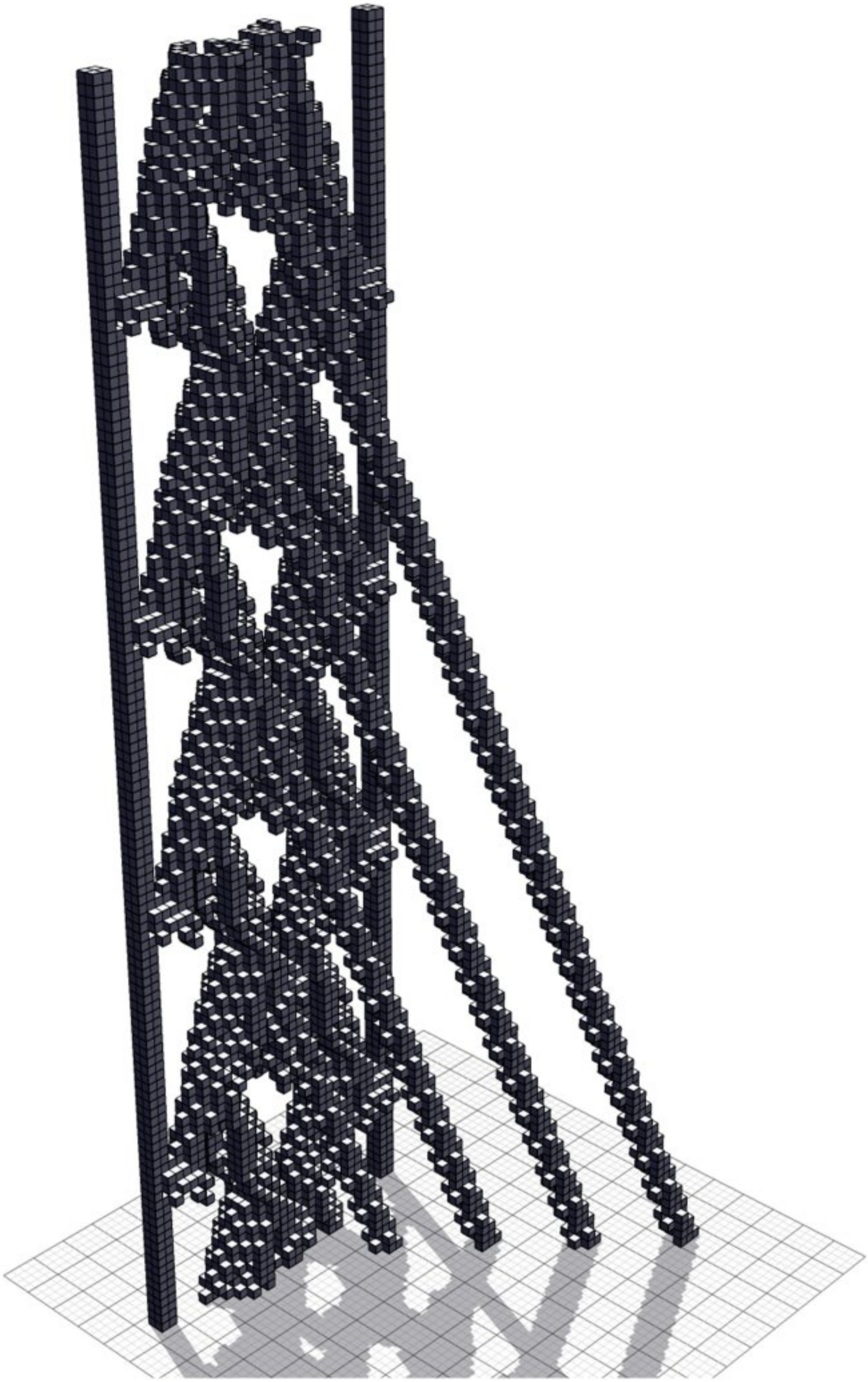
Ka Lok Man
Xi'an Jiaotong-Liverpool University
204 PUBLICATIONS **1,513** CITATIONS
SEE PROFILE

Jeremy Smith
University of Liverpool
146 PUBLICATIONS **2,347** CITATIONS
SEE PROFILE

IJDATICS

ISSN 2071-2987

International Journal of Design, Analysis and Tools for Integrated Circuits and Systems



Preface

Welcome to the Volume 11 Number 1 of the International Journal of Design, Analysis and Tools for Integrated Circuits and Systems (IJDATICS). This volume is comprised of selected research papers from the International Conference on Recent Advancements in Computing in Artificial Intelligence, Internet of Things and Computer Engineering Technology (CICET), October 24-26, 2022, Taipei, Taiwan. CICET 2022 is hosted by The Tamkang University amid pleasant surroundings in Taipei, which is a delightful city for the conference and traveling around.

CICET 2022 serves a communication platform for researchers and practitioners both from academia and industry in the areas of Computing in Artificial Intelligence (AI), Internet of Things (IoT), Integrated Circuits and Systems and Computer Engineering Technology. The main target of CICET 2022 is to bring together software/hardware engineering researchers, computer scientists, practitioners and people from industry and business to exchange theories, ideas, techniques and experiences related to all aspects of CICET. Recent progress in Deep Learning (DL) has unleashed some of the promises of AI, moving it from the realm of toy applications to a powerful tool that can be leveraged across a wide number of industries. In recognition of this, CICET 2022 has selected Artificial AI and Machine Learning (ML) as this year's central theme.

The Program Committee of CICET 2022 consists of more than 150 experts in the related fields of CICET both from academia and industry. CICET 2022 is organized by The Tamkang University, Taipei, Taiwan, and co-organized by AI University Research Centre (AI-URC) and Research Institute of Big Data Analytics (RIBDA), Xi'an Jiaotong-Liverpool University, China as well as supporting by: Swinburne University of Technology Sarawak Campus, Malaysia; Taiwanese Association for Artificial Intelligence, Taiwan; Trcuteco, Belgium; International Journal of Design, Analysis and Tools for Integrated Circuits and Systems, International DATICS Research Group. The CICET 2022 Technical Program includes 1 invited speaker and 30 oral presentations. We are beholden to all of the authors and speakers for their contributions to CICET 2022. On behalf of the program committee, we would like to welcome the delegates and their guests to CICET 2022. We hope that the delegates and guests will enjoy the conference.

Professor Ka Lok Man, Xi'an Jiaotong-Liverpool University, China

Professor Young B. Park, Dankook University, Korea

Chairs of CICET 2022

CICET

Table of Contents

Vol. 11, No. 1, November 2022

E-mail text deception detection based on Machine Learning technology

Hongjian Zhang *, and Gabriela Mogos

Abstract— **In January 2022, the number of global Internet users will reach 4.95 billion, and Internet users account for 62.5% of the total population [1]. As the number of users grows, the content on the Internet expands by the minute.**

At the same time, e-mail is increasingly used, with more than a third of the world's population now using it [7]. Malicious people can use e-mail to commit fraud, and users often suffer losses if they are unprepared. So, the motivation for this paper was to explore what techniques could be used to reduce the amount of email fraud and prevent email users from suffering financial or personal information loss.

Index Terms— **Machine Learning, NLP, Email deception detection.**

I. INTRODUCTION

Machine learning is adaptive, that is, the system will use the accumulation of data, automatic learning, and training to improve system performance. Machine learning techniques are developed from statistics and optimization theory. Up to now, many different algorithms have been developed, such as Logistic Regression (LR), support vector machine, decision tree, Naive Bayes and some other algorithms, which are important ways to data analysis and mining problems.

Logistic regression is very easy to use and can be used in many scenarios. It is especially suitable for the analysis of dichotomies and disordered nominal multivariate dependent variables. For ordinal multivariate dependent variables, multivariate logistic regression analysis can also be considered, but in some other models, including weighted least squares and linear regression, are related to multivariate and need to be considered when using them [2].

In 1964, Support Vector Machine (SVM) technology was already in its infancy. And after 1990, rapid development and derivation of many improved algorithms, these achievements have been applied in a wide range of fields. For example, SVM can learn by examining a large number of credit card activities and can identify whether a credit card activity contains fraudulent intent after training whether these activities are fraudulent or not. Alternatively, SVM recognizes handwritten digits by analyzing a large number of handwritten digital images and scanning them [5].

Decision tree is actually an analysis method with a long history. Now, decision tree is used in machine learning to replace "human" experience with the principles of mathematics and statistics, so that the machine can automatically generate judgment logic from data [4].

Before the technology of machine learning, the theoretical basis for Naive Bayes was introduced by the British mathematician Thomas Bayes. He argued that when you don't know exactly what a thing is, you can judge the probability of its essential properties by the number of events related to its particular nature. Naive Bayes performs well in complex environments compared to other classifiers. And it applies to data with independent dimensions [6].

In fact, there are many more machine learning algorithms, and each algorithm has a different effect in a particular scene. Therefore, in practical application, more of the same group of data is applied to multiple models for training and testing, and then compared.

The purpose of this paper is to use machine learning techniques to explore which models might be suited for predicting which parts of emails are more likely to be spam. The trained model can be used to predict a wider range of emails and timely alert users if the results are likely to be fraudulent emails.

There are two main technologies in this paper. Firstly, the Natural Language Process (NLP) of e-mail text is carried out, and more information dimensions are obtained after processing the text. Then, various machine learning models are used to train and test in these dimensions, and then the prediction results are compared.

This research considers that some e-mails are accompanied by certain words, and these words contain certain tendencies from the author of the e-mail, so some specific words are found through classification. These words form a cloud map that users can view to see if the email they receive is fraudulent.

II. METHODOLOGY

In order to find a model that is more suitable for detecting spams, the same data is used here to find out the model with higher score. Data preprocessing is carried out at first, and then several models are trained and tested to get scores for comparison.

A. Data processing

It can be found that there are a lot of symbols in the *Message_body* data like “*”, “@” or “&”. These symbols are

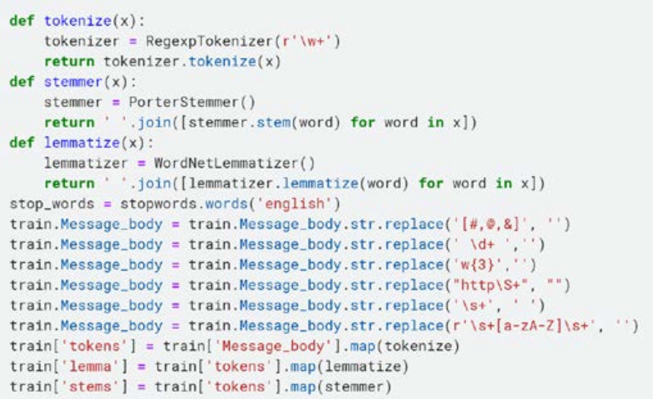
All authors are with the Department of Computing, School of Advanced Technology, Xi’an Jiaotong-Liverpool University, Suzhou, China. (email: Gabriela.Mogos@xjtlu.edu.cn).

of limited use in prediction, so they are cleared in the pre-processing stage.

Meanwhile, *url* links and numbers in the email text were found to have a negative impact on the accuracy of the prediction after several sessions of training. So, this information is also removed from the *Message_body* during the preprocessing phase.

Once the symbols and content are removed, the word segmentation of the text is simplified using the *RegexpTokenizer*. Then use *WordNetLemmatizer* to convert synonyms to make the model more general. Finally, use *PorterStemmer* to make the text more standard.

The preprocessing code for train data is shown like figure 1.

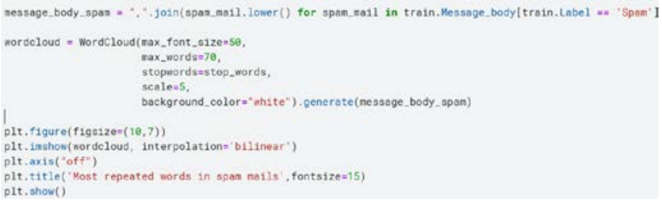


```
def tokenize(x):
    tokenizer = RegexpTokenizer(r'\w+')
    return tokenizer.tokenize(x)
def stemmer(x):
    stemmer = PorterStemmer()
    return ' '.join([stemmer.stem(word) for word in x])
def lemmatize(x):
    lemmatizer = WordNetLemmatizer()
    return ' '.join([lemmatizer.lemmatize(word) for word in x])
stop_words = stopwords.words('english')
train.Message_body = train.Message_body.str.replace('[#,@,&]', '')
train.Message_body = train.Message_body.str.replace(' \d+ ','')
train.Message_body = train.Message_body.str.replace('w{3}','')
train.Message_body = train.Message_body.str.replace("http\S+", "")
train.Message_body = train.Message_body.str.replace('\s+', ' ')
train.Message_body = train.Message_body.str.replace(r'\s+[a-zA-Z]\s+', '')
train['tokens'] = train['Message_body'].map(tokenize)
train['lemma'] = train['tokens'].map(lemmatize)
train['stems'] = train['tokens'].map(stemmer)
```

Fig. 1. The data preprocessing code.

B. Cloud

The occurrence of certain words is high frequency, and word clouds can be formed according to these high frequency words. This can be used to give the observer a sense of which terms are most frequently used in spam, and which are most frequently used in non-spam.



```
message_body_spam = ",".join(spam_mail.lower() for spam_mail in train.Message_body[train.Label == 'Spam']

wordcloud = WordCloud(max_font_size=50,
                      max_words=70,
                      stopwords=stop_words,
                      scale=5,
                      background_color="white").generate(message_body_spam)

plt.figure(figsize=(10,7))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.title('Most repeated words in spam mails',fontsize=15)
plt.show()
```

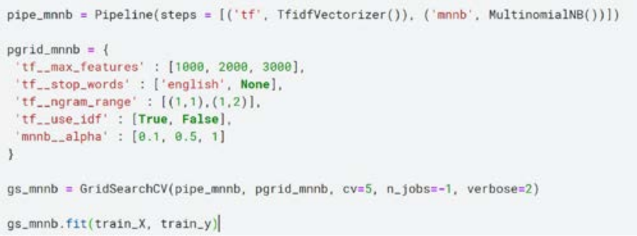
Fig. 2. The word cloud code

C. Naïve Bayes

Naive Bayes is an approach based on Bayes' theorem and the assumption of feature condition independence. Assume that the attributes are conditionally independent of each other when the target value is given [9]. Multinomial Naive Bayes MNB is used in the project.

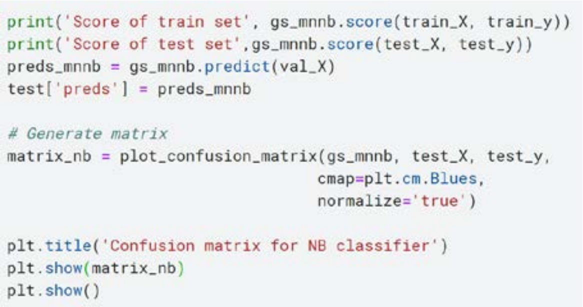
The MNB function is used to find parameters suitable for this data. We first used *GridSearchCV* function to adjust parameters automatically and found parameters more suitable for this data, including *max_features*, *ngram_range* and so on.

These parameters are then used to train the data.



```
pipe_mnnb = Pipeline(steps = [('tf', TfidfVectorizer()), ('mnnb', MultinomialNB())])

pgrid_mnnb = {
 'tf__max_features' : [1000, 2000, 3000],
 'tf__stop_words' : ['english', None],
 'tf__ngram_range' : [(1,1),(1,2)],
 'tf__use_idf' : [True, False],
 'mnnb__alpha' : [0.1, 0.5, 1]
}

gs_mnnb = GridSearchCV(pipe_mnnb, pgrid_mnnb, cv=5, n_jobs=-1, verbose=2)

gs_mnnb.fit(train_X, train_y)
```

Fig.3. GridSearchCV function of MNB



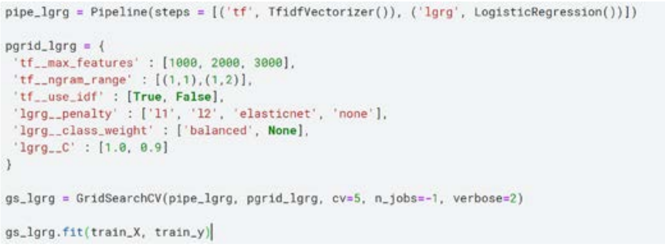
```
print('Score of train set', gs_mnnb.score(train_X, train_y))
print('Score of test set',gs_mnnb.score(test_X, test_y))
preds_mnnb = gs_mnnb.predict(val_X)
test['preds'] = preds_mnnb

# Generate matrix
matrix_nb = plot_confusion_matrix(gs_mnnb, test_X, test_y,
                                  cmap=plt.cm.Blues,
                                  normalize='true')

plt.title('Confusion matrix for NB classifier')
plt.show(matrix_nb)
plt.show()
```

Fig.4. MNB code

D. Logistic Regression

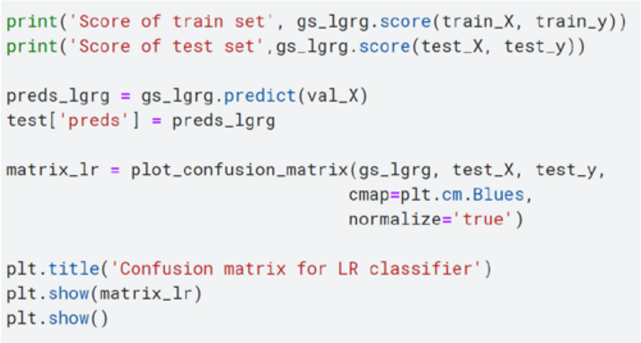
Through the Logistic function, whether the data is spam mapped to a probability value between 0 and 1, and the classification of the data can be obtained by comparing with 0.5 [3].

In the application of Logistic Regression (LR) algorithm, penalty term, regularization coefficient, weight and other parameters are considered to ensure the accuracy of prediction. Similarly, the *GridSearchCV* function was used to determine the parameters and find the appropriate parameters.



```
pipe_lgrg = Pipeline(steps = [('tf', TfidfVectorizer()), ('lgrg', LogisticRegression())])

pgrid_lgrg = {
 'tf__max_features' : [1000, 2000, 3000],
 'tf__ngram_range' : [(1,1),(1,2)],
 'tf__use_idf' : [True, False],
 'lgrg__penalty' : ['l1', 'l2', 'elasticnet', 'none'],
 'lgrg__class_weight' : ['balanced', None],
 'lgrg__C' : [1.0, 0.9]
}

gs_lgrg = GridSearchCV(pipe_lgrg, pgrid_lgrg, cv=5, n_jobs=-1, verbose=2)

gs_lgrg.fit(train_X, train_y)
```

Fig.5. *GridSearchCV* function of Logistic Regression LR

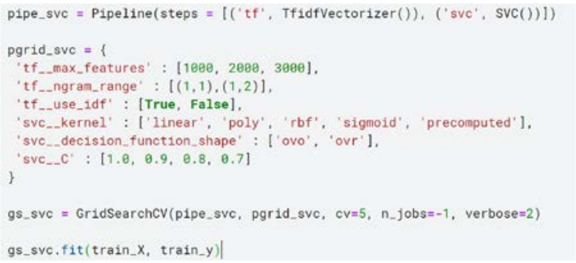


```
print('Score of train set', gs_lgrg.score(train_X, train_y))
print('Score of test set',gs_lgrg.score(test_X, test_y))

preds_lgrg = gs_lgrg.predict(val_X)
test['preds'] = preds_lgrg

matrix_lr = plot_confusion_matrix(gs_lgrg, test_X, test_y,
                                  cmap=plt.cm.Blues,
                                  normalize='true')

plt.title('Confusion matrix for LR classifier')
plt.show(matrix_lr)
plt.show()
```

Fig. 6. The LR code

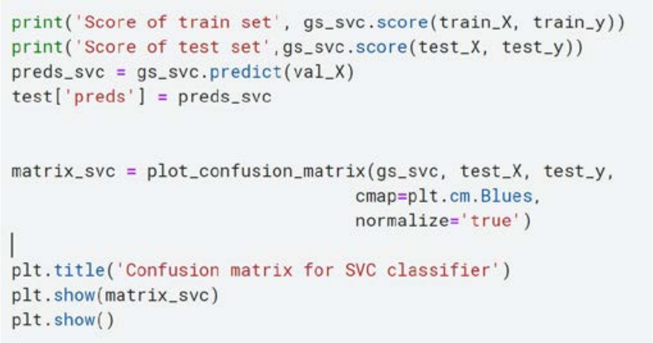
E. *Support Vector Classification*

Support Vector Classification SVM is a supervised machine learning algorithm that can be used for classification or regression challenges [8]. In this algorithm, each data item is regarded as a point in n-dimensional space as a point, and each eigenvalue is the value of a specific coordinate. Since support vector machine cannot tolerate non-standard data well, the data is carefully cleaned during data preprocessing to ensure the accuracy of SVC. *GridSearchCV* are also used to find suitable parameter values.



```
pipe_svc = Pipeline(steps = [('tf', TfidfVectorizer()), ('svc', SVC())])

pgrid_svc = {
 'tf__max_features' : [1000, 2000, 3000],
 'tf__ngram_range' : [(1,1),(1,2)],
 'tf__use_idf' : [True, False],
 'svc__kernel' : ['linear', 'poly', 'rbf', 'sigmoid', 'precomputed'],
 'svc__decision_function_shape' : ['ovo', 'ovr'],
 'svc__C' : [1.0, 0.9, 0.8, 0.7]
}

gs_svc = GridSearchCV(pipe_svc, pgrid_svc, cv=5, n_jobs=-1, verbose=2)

gs_svc.fit(train_X, train_y)
```

Fig.7. The *GridSearchCV* function of SVC



```
print('Score of train set', gs_svc.score(train_X, train_y))
print('Score of test set',gs_svc.score(test_X, test_y))
preds_svc = gs_svc.predict(val_X)
test['preds'] = preds_svc


matrix_svc = plot_confusion_matrix(gs_svc, test_X, test_y,
                                   cmap=plt.cm.Blues,
                                   normalize='true')

plt.title('Confusion matrix for SVC classifier')
plt.show(matrix_svc)
plt.show()
```

Fig.8. The SVC code

III. RESULTS

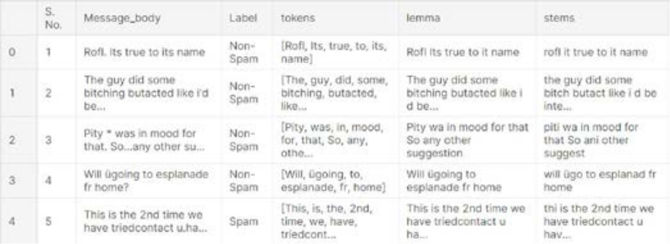
A. *Data processing*

Before data preprocessing, the downloaded data contains three attributes as shown in the following table 1: serial number, message text and label.

After removing *symbols*, *numbers*, *URL*, and so on, and dividing words, the original table looks like the following Figure 9.

S. No.	Message_body	Label
1	Rofl. It's true to its name	Non-Spam
2	The guy did some bitching, but we acted like we'd be interested in buying something else next week and he gave it to us for free	Non-Spam
3	Pity, * was in mood for that. So... any other suggestions?	Non-Spam
4	Will ?b going to esplanade fr home?	Non-Spam
5	This is the 2nd time we have tried 2 contact u. U have won the ?50 Pound prize. 2 claim is easy, call 087187272008 NOW1! Only 10p per minute. BT-national-rate.	Spam

Table 1. The original table



	S. No.	Message_body	Label	tokens	lemma	stems
0	1	Rofl. Its true to its name	Non-Spam	[Rofl, Its, true, to, its, name]	Rofl Its true to it name	rofl it true to it name
1	2	The guy did some bitching butacted like i'd be...	Non-Spam	[The, guy, did, some, bitching, butacted, like...	The guy did some bitching butacted like i d be...	the guy did some bitch butact like i d be inte...
2	3	Pity * was in mood for that. So...any other su...	Non-Spam	[Pity, was, in, mood, for, that, So, any, othe...	Pity wa in mood for that So any other suggestion	piti wa in mood for that So ani other suggest
3	4	Will ūgoing to esplanade fr home?	Non-Spam	[Will, ūgoing, to, esplanade, fr, home]	Will ūgoing to esplanade fr home	will ūgo to esplanad fr home
4	5	This is the 2nd time we have triedcontact u.ha...	Spam	[This, is, the, 2nd, time, we, have, triedcont...	This is the 2nd time we have triedcontact u ha...	thi is the 2nd time we have triedcontact u hav...

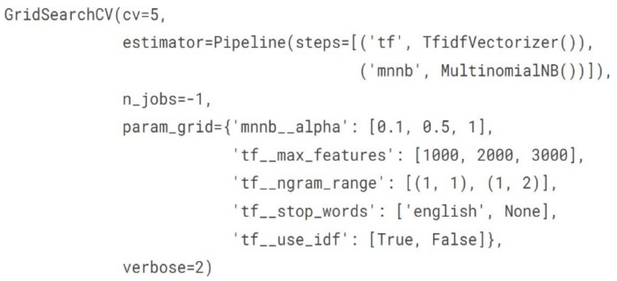
Fig.9. Data processing table

B. *Word Cloud*

In the two resulting word cloud images, we can observe some very clear similarities and differences. Common verbs like *get*, *call*, and *see* are very common in both Spam and non-spam, as are time words like today and time

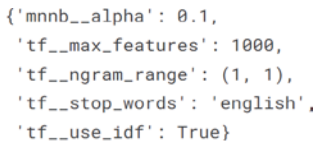
In non-Spam, some words have subjective feelings such as *love* and *like*, while in Spam, there is no such expression. The most frequently used words in spam are *cash, service, please*, and so on, all of which indicate an attempt by the author to elicit a response from the recipient.

C. Naive Bayes



```
GridSearchCV(cv=5,
             estimator=Pipeline(steps=[('tf', TfidfVectorizer()),
                                       ('mnnb', MultinomialNB())]),
             n_jobs=-1,
             param_grid={'mnnb__alpha': [0.1, 0.5, 1],
                         'tf__max_features': [1000, 2000, 3000],
                         'tf__ngram_range': [(1, 1), (1, 2)],
                         'tf__stop_words': ['english', None],
                         'tf__use_idf': [True, False]},
             verbose=2)
```

Fig.10. MNB GridSearchCV parameters

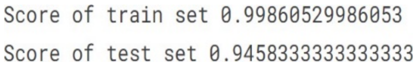


```
{'mnnb__alpha': 0.1,
 'tf__max_features': 1000,
 'tf__ngram_range': (1, 1),
 'tf__stop_words': 'english',
 'tf__use_idf': True}
```

Fig. 11. The best parameters

In MNB training and testing, the above figures were finally obtained. Several parameters were most suitable for this data to be trained using MNB model.

The training set's score is close to 1; while the test set's score is 0.946. According to figure 12, it can be found where the error occurred. The accuracy was 0.98 for non-spam prediction, but only 0.72 for spam prediction. This deviation is large.



```
Score of train set 0.99860529986053
Score of test set 0.9458333333333333
```

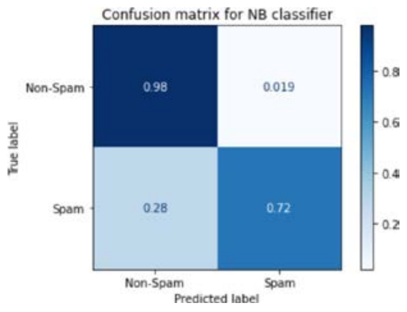
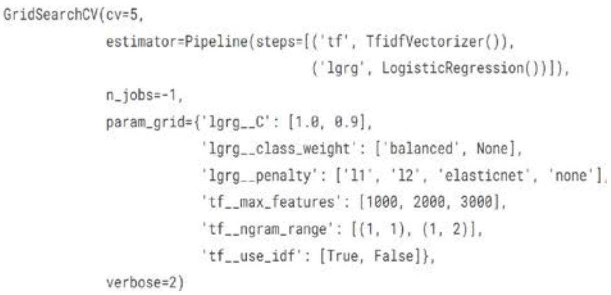


Fig.12. MNB training and testing

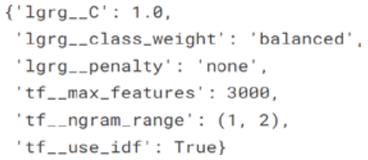
D. Logistic Regression

Figure 14 shows most suitable parameters for this data to be trained using LR model. The training set's score is 1; while the test set's score is 0.967. According to figure 15, it can be found where the error occurred. The accuracy was very close to 1 for non-spam prediction, and 0.78 for spam prediction. This deviation is still large.



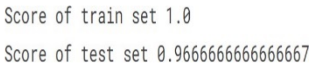
```
GridSearchCV(cv=5,
             estimator=Pipeline(steps=[('tf', TfidfVectorizer()),
                                       ('lgrg', LogisticRegression())]),
             n_jobs=-1,
             param_grid={'lgrg__C': [1.0, 0.9],
                         'lgrg__class_weight': ['balanced', None],
                         'lgrg__penalty': ['l1', 'l2', 'elasticnet', 'none'],
                         'tf__max_features': [1000, 2000, 3000],
                         'tf__ngram_range': [(1, 1), (1, 2)],
                         'tf__use_idf': [True, False]},
             verbose=2)
```

Fig.13. LR GridSearchCV parameters



```
{'lgrg__C': 1.0,
 'lgrg__class_weight': 'balanced',
 'lgrg__penalty': 'none',
 'tf__max_features': 3000,
 'tf__ngram_range': (1, 2),
 'tf__use_idf': True}
```

Fig. 14. The best parameters



```
Score of train set 1.0
Score of test set 0.9666666666666667
```

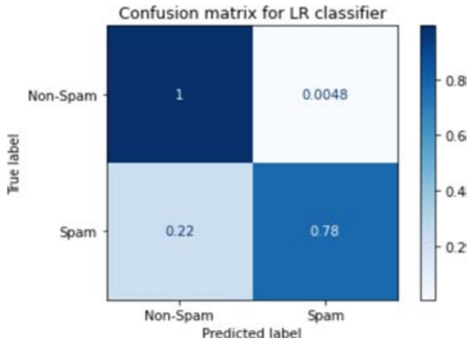
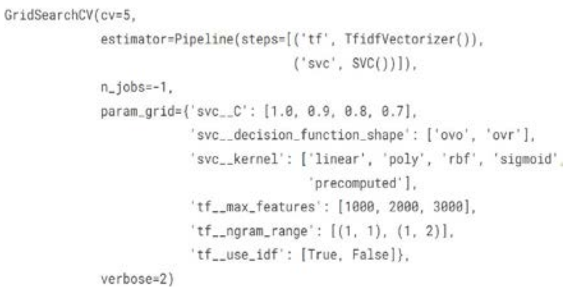


Fig.15. LR training and testing

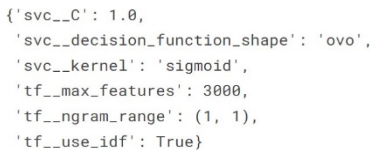
E. Support Vector Classification

The training set's score is 0.994; while the test set's score is 0.9625. According to figure 18, it can be found where the error occurred. The accuracy was very close to 1 for non-spam prediction, and 0.75 for spam prediction. This deviation is still large.



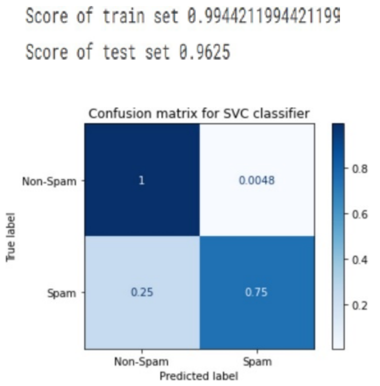
```
GridSearchCV(cv=5,
             estimator=Pipeline(steps=[('tf', TfidfVectorizer()),
                                       ('svc', SVC())]),
             n_jobs=-1,
             param_grid={'svc__C': [1.0, 0.9, 0.8, 0.7],
                         'svc__decision_function_shape': ['ovo', 'ovr'],
                         'svc__kernel': ['linear', 'poly', 'rbf', 'sigmoid',
                                         'precomputed'],
                         'tf__max_features': [1000, 2000, 3000],
                         'tf__ngram_range': [(1, 1), (1, 2)],
                         'tf__use_idf': [True, False]},
             verbose=2)
```

Fig.16. SVC GridSearchCV parameters



```
{'svc__C': 1.0,
 'svc__decision_function_shape': 'ovo',
 'svc__kernel': 'sigmoid',
 'tf__max_features': 3000,
 'tf__ngram_range': (1, 1),
 'tf__use_idf': True}
```

Fig. 17. The best parameters



Score of train set 0.9944211994421199

Score of test set 0.9625

Fig.18. SVC training and testing

E. *Comparison*

The test scores of the three algorithms were 0.946, 0.967 and 0.9625 respectively. The accuracy of spam prediction was 0.72, 0.78, 0.75 respectively. Therefore, for this data set, the Logistic Regression model performed best in training and testing.

IV. CONCLUSIONS

The data set used in this project is a fraction of the mails generated on a daily basis. In terms of data, due to the difficulty in finding Chinese email message data sets, English email data sets were selected at last. The data were just processed into three dimensions which are tokens, lemma, and stems. Although it has strong universality, it may need to divide more dimensions for testing in a large amount of data to improve accuracy.

More dimensions are added for training, including the word count and title of email, whether to carry attachments, the number of URL links and so on, and the accuracy of fitting may be higher.

REFERENCES

[1] Ben. (2019). Do you know how many emails are sent and received around the world every day? Available at: https://zhuanlan.zhihu.com/p/76152504. (Accessed: 2 May 2022).

[2] Menard, S. (2002). Applied logistic regression analysis (Vol. 106).

[3] Menard, S. W. (2010). Logistic regression: from introductory to advanced concepts and applications. SAGE. Available at: https://search-ebscohost-com.ez.xjtlu.edu.cn/login.aspx?direct=true&db=cat01010a&AN=xjtlu.0000805129& site=eds-live&scope=site (Accessed: 2 May 2022).

[4] Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. Journal of Chemometrics: A Journal of the Chemometrics Society, 18(6), 275-285.

[5] Noble, W. S. (2006). What is a support vector machine?. Nature biotechnology, 24(12), 1565-1567.

[6] Rish, I. (2001). An empirical study of the naive Bayes classifier, IJCAI 2001 workshop on empirical methods in artificial intelligence, 3(22), 41-46.

[7] Xiaohong.Guan. (2022). Analysis of the number of Internet users, proportion of Internet users, online duration and reasons. Available at: chyxx.com/industry/1106494.html. (Accessed: 6 May)

[8] Yunqian Ma and Guodong Guo (2014). Support Vector Machines Applications. Cham: Springer. Available at: https://search.ebscohost.com/login.aspx?direct=true&db=edsebk&AN=699741&site= eds-live&scope=site (Accessed: 2 May 2022).

[9] Yuslee, N. S. and Abdullah, N. A. S. (2021). 'Fake News Detection using Naive Bayes', 2021 IEEE 11th International Conference on System Engineering and Technology (ICSET), doi: 10.1109/ICSET53708.2021.9612540.

ReSoNate: A Protocol for Audio Transmission over Low Power Wide Area Networks

Shuaibu Musa Adam, Yandi Liu, Absar-Ul-Haque Ahmar, Sam Michiels and Danny Hughes

***Abstract*—Low Power Wide Area Networks (LPWANs), such as LoRa, enable end-users to create low power networks that cover 10s of km with a single gateway, providing low cost connectivity to areas that may be poorly served by the mainstream cellular networks. However, the low data rates of current LPWANs have limited their applicability to plain text, sensor and control applications. This paper explores whether extremely low bitrate audio codecs can deliver adequate quality real-time voice communication over LPWANs while preserving low power operation. Specifically, we contribute ReSoNate, an efficient half-duplex voice communication protocol for LoRa that builds on CODEC 2. We created a reference implementation of ReSoNate for a representative embedded platform (100MHz ARM Cortex-M4 with 128kB of RAM and 512kB of Flash) and tested it with the RFM9x LoRa transceiver. Energy consumption and audio quality assessments were then conducted to investigate its performance. Our results show that: (i.) ReSoNate achieves acceptable audio quality for basic voice communication, (ii.) the energy profile of the reference implementation can achieve long battery lifetimes in realistic settings (iii.) the protocol is robust to high levels of packet loss of up to 20%. Considered in sum, the contributions of this paper pave the way for the deployment of extremely low cost and low power voice communication networks in remote areas such as the developing world.**

***Index Terms*— LoRa, Voice communication, Internet of Things, Low-Power Wide-Area Network (LPWAN).**

I. INTRODUCTION

Low Power Wide Area Network technologies (LPWAN) enable the Internet-of-Things (IoT) to benefit from battery-powered networks offering wide area coverage at a low-cost for low bit rate traffic [10]. LPWAN technologies include licensed or license-free variants. If security, reliability and high-speed communications are the priorities, then licensed band solutions are typically preferred, which include: Narrowband-IoT (NB-IoT), Extended Coverage Global System for Mobile Communications (EC-GSM), and Long Term Evolution for Machines (LTE-M). However, if low cost is prioritised, then Sigfox and LoRaWAN, which operate in the license-free frequency bands are more suitable [19].

LoRa networks, for example, are employed in healthcare [20, 21], localisation [6], precision agriculture [16, 17], sailing [8], and smart cities [1,18]. However, despite its potential, LoRaWAN technology is strictly regulated to a typical duty cycle of 1% and 14 dBm transmission power [5], resulting in maximum data rates of a few kbps. Nevertheless, several studies have attempted to use LoRa to transmit images [13, 14, 15], voice [9, 12] or both [7]. As yet however, no work has managed to achieve live audio transmission within the EU frequency band limitations of LoRa.

In this paper, we propose ReSoNate, a half-duplex real-time audio protocol and associated reference implementation for LoRaWAN. Initial results show that ReSoNate

1) Achieves live audio transmission within the frequency bands limit of the EU regulations (i.e. 1% duty cycle), by using the Codec 2 audio encoder in 1.3 kbps mode [11].
2) Offers reasonable audio quality even with a packet loss ratio of up to 20%, as confirmed by a small-scale study.
3) Supports audio communication on a pair of 2400mAh AA LiSO2 batteries for multiple days on a single charge.

The ReSoNate prototype confirms the feasibility of wireless audio over LoRa with a very low-rate audio codec using simple hardware components. The software code and design are available in open-source, enabling interested parties to further extend and improve the current prototype.(GitHub)

The remainder of this paper is structured as follows. Section II describes the design of ReSoNate. Section III provides important implementation details. Section IV describes our experiments using the reference platform to evaluate the performance of ReSoNate. Section V reviews related work. Finally, Section VI concludes and discusses future work.

II. DESIGN

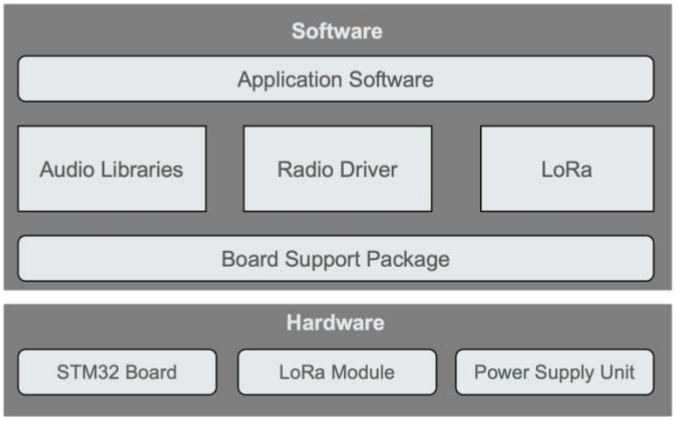


Fig. 1. Simplified software-hardware architecture of ReSoNate

A. Reference Hardware

The STM32F411E Discovery kit (F411E board) [4] is based on the STM32F411VET6 [23], an ARM-Cortex M4 CPU with a single-precision floating-point unit (FPU) running at a maximum clock frequency of 100 MHz. It integrates 512 Kbytes Flash memory and 128 Kbytes SRAM with a Direct

All authors are with the imec-DistriNet, KU Leuven, B-3001 Leuven, Belgium. email: {firstname.lastname}@ kuleuven.be

Memory Access (DMA) controller to manage the memory-peripheral transfers. The F411E board has an onboard microphone for audio input and an audio output jack for playback. There are also four programmable LEDs of different colours as well as *reset* and *user buttons*, respectively.

The microphone generates digital audio in *Pulse-Density Modulation* (PDM) format, while both the Codec 2 coder and the audio DAC require *Pulse-Code Modulation* (PCM). As a result, conversion is required from PDM to PCM. Moreso, one microphone only generates one channel of audio, or mono sound, which works well for the coder, but the audio DAC needs stereo audio input. The solution is to duplicate the mono audio and feed it to both the left and right channels to make the stereo audio.

The LoRa transceiver module consists of an Adafruit RFM9x radio module and a monopole antenna. RFM9x is based on the SEMTECH SX1276 LoRa module, which, in Europe, operates at 868 MHz. The module is connected to the dev-board using SPI. The low-power characteristics of STM32F411E board and the LoRa module enable long battery life. The reference hardware design is shown in Figure 2.

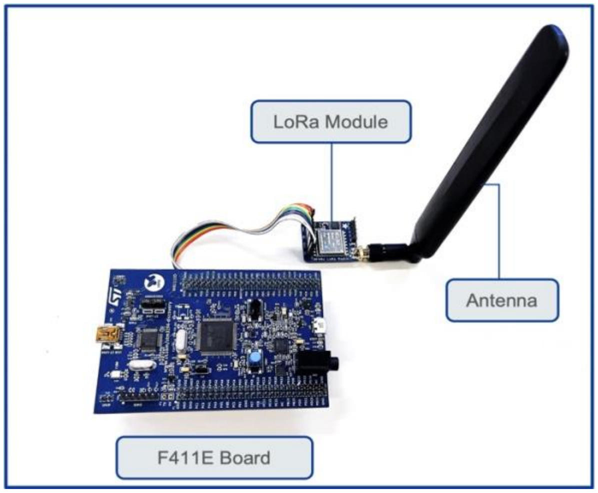


Fig. 2. ReSoNate hardware design

B. Software Stack

1. Audio Libraries: The Codec 2 libraries used in this research are a modified version of the official implementation [11] that is extended to avoid the use of double-precision floating-point numbers, hence increasing efficiency on low-end embedded computing platforms that lack the required hardware.
2. Board Support Package: ReSoNate uses CMSIS-CORE to initialise the system and access standard registers, while the STM32F4 HAL library provides generic functions, such as configuring peripherals and handling interrupts. The CMSIS-DSP library provides the core mathematics functions used by codec2. Finally, the PDM2PCM library, is used to convert mono PDM format audio to stereo PCM format audio as required by codec2. Standard drivers are used for the onboard microphone and audio DAC.

3. Radio Driver: To implement the radio driver, the STM32 HAL driver for the LoRa SX1278 module [3] is used with a small modification to accommodate generated interface code from the STM32 development environment. As the driver uses the STM32 HAL interfaces, it can be conveniently migrated to other STM32-based platforms.

C. End-to-End Data Flow

The flow of speech data from transmitter to the receiver is illustrated in Figure 3. On the transmitter side, the human voice first goes through the microphone to the Analog-to-Digital Converter (ADC), where it becomes digital signals. The signal is then converted and processed by the Codec 2 encoder into binary content named *c2bits*. The LoRa transceiver sends out the data as a sequence of standard LoRaWAN packets. After the remote device receives the *c2bits*, it decodes the data. Finally, the signal goes through the DAC, which may be attached to a speaker or headphones to be heard by the listener.

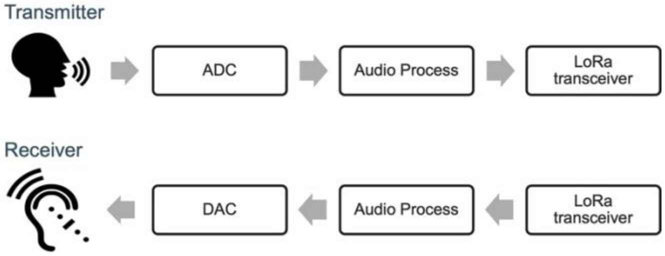


Fig. 3. End-to-end data flow for ReSoNate

III. IMPLEMENTATION

A. Board Connection

A total of four serial interfaces are enabled on the F411E board. First, the SPI1 interface uses the PA5, PA6 and PA7 pins to communicate with the LoRa module. Second, the I2S2 interface employs the pins PB10 and PC3 to communicate with the onboard microphone. Third, the pins PA4, PC7, PC10, and PC12 are controlled by the I2S3 interface to communicate with the audio DAC. Lastly, the USART1 interface operates the pins PA15 and PB3 to communicate with a PC. Table 1 shows the wiring between the F411E board and the LoRa module.

Table 1. Wiring between the F411E board and the LoRa module

F411E board pins	RFM9x LoRa module pins
GND	GND
3V	VIN
PA2	GO
PA5	SCK
PA6	MISO
PA7	MOSI
PA10	CS
PC9	RST

The *user button* binds to pin PA0 and is configured to trigger interrupts when it is pressed or released. A variable *UserPressButton* tracks the state of the button. When a user presses the button, a rising edge interrupt occurs in PA0, and *UserPressButton* is set to 1. When the user releases the button, a falling edge interrupt is triggered, and *UserPressButton* is assigned to 0. The user button is programmed as a push-to-talk button by checking the *UserPressButton* value.

B. *Application State Machine*

The application can be divided into four states: *(i.) recording, (ii.) transmission, (iii.) receiving,* and *(iv.) playback*.

(i.) Recording: When a user *presses and holds* the *user button*, the application enters the *recording state*, during which the input audio is converted from PDM to PCM and stored in an array in RAM. The device then enters the *transmission state.*

(ii.) Transmission: The board begins transmitting the c2bits once the array is full, or the user *releases* the record button. This continues until all data has been transmitted.

(iii.) Receiving: The application remains in the receiving state until it receives a packet. Once a packet is received, the payload is stored in the same array to minimise memory consumption and playback is triggered.

(iv.) Playback: The received *c2bits* are *decoded and played*, after which the application returns to the receiving state.

The size of the encoded array is configurable and determines the maximum duration of the recording. In the current implementation, the size is configured to 2100 bytes and corresponds to a total duration of 12 seconds.

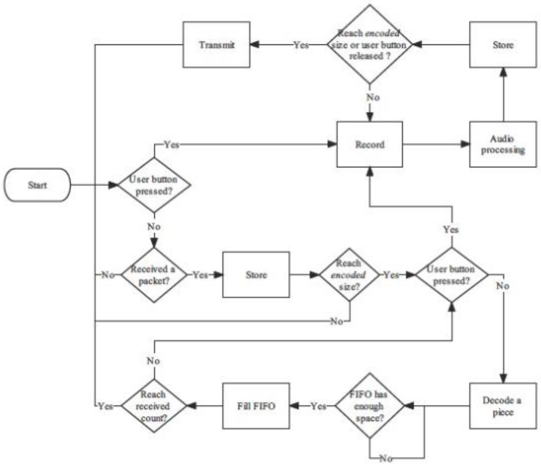


Fig. 4. Semi-live audio

IV. EVALUATION

We designed a series of experiments to test the energy consumption and performance of ReSoNate. These are reported in Section IV.A and IV.B respectively.

A. *Audio Energy Consumption*

We quantified the energy consumption of ReSoNate in each phase of its operation (receiving state, recording state, and finally transmission & playback state) on the F411E evaluation board. All tests were performed at 3V. To ensure accurate and consistent measurements all tests were carried out 10 times and averaged.

Receiving State: Immediately, after the device is powered ON it enters the *receiving state*. In this state, it consumes an average of 25.0mA.

Recording State: Fluctuations in the energy consumption between receiving and recording states are negligible. Recording state energy consumption was measured in three stages: start-of-recording values, peak recording values and end-of-recording values, respectively. Average values are found to be 26.86mA, 26.93mA and 32.90mA respectively. The total recording time ranged from 10 to 20 seconds; with the sample energy measurements at the interval of 500 ms.

Transmission & Playback State: The transmission and playback states could have been measured separately, but the experiments were constrained by measuring the combined parameters. This state is measured immediately after the recording stopped, and the *user button* is released. Interestingly, in this state, it was observed that the energy consumption decreases less than the receiving and recording states with average values of around 20.0mA. After which the energy consumption increases with an average peak value of around 31.52mA (which is still below the average peak value of the recording state). Finally, the energy consumption decreases with a linear value until the last playback point.

Table 2 estimates the battery life of ReSoNate when using a pair of standard 3.6V 2400mAh LiSO2 batteries (for 4800mAh total) in each of these phases of operation:

Table 2. Estimated battery lifetime

Phase of operation	Battery lifetime
Receiving	8 days
Recording	6.1 days
Transmission/Playback	6.4 days

As can be seen from Table 2, ReSoNate delivers extremely long talk-times using a single battery charge. However, further improvements are still possible by using techniques such as time-synchronisation to reduce the power costs of waiting for an incoming call. In our future work, we will explore how this can be accomplished by building on our prior work [22, 24].

B. *Audio Quality Test*

In this section, we first analyse whether the audio quality offered by ReSoNate running on the reference platform compares to the standard Codec 2 implementation running on a mainstream PC. We then investigate the resilience of ReSoNate to packet loss and thereby its robustness.

1) Audio Quality under Different Conditions: In this test, three variables are controlled as shown in Table 3: the microphone, the platform running Codec 2, and the playback hardware. The microphone is either a smartphone microphone ("*external*") or the F411E board microphone ("*STM*"). The Codec 2 runs on the PC or the F411E board. The audio playback is either on the PC or by the F411E onboard audio DAC. The smartphone used is a Redmi K20 Pro, and the PC has a four-core CPU running at 2.6 GHz and 20 GB RAM. The PC Codec 2 implementation operates in a virtual machine with a Linux operating system.

A listening test is conducted for the assessment, presented in the form of a questionnaire created using Google Forms. In the test, assessors first listen to a *reference speech audio*, which is recorded by the smartphone and down-sampled to normalised loudness of 8 kHz. It also serves as input audio for *conditions 1-3*. The assessors then listen to six audio clips, each processed under the respective conditions shown in the table. All the clips have the same text content in English voiced by a single person. Assessors rate each audio clip by giving them a score related to its quality. There are six options, numbers 1 to 6, for the rating, with 1 indicating worst and 6 indicating best quality.

Table 3. Conditions of the processed audio

Condition	Microphone	Codec 2 platform	Playback hardware
1	External	PC	PC
2	External	STM	PC
3	External	STM	STM
4	STM	PC	PC
5	STM	STM	PC
6	STM	STM	STM

A total of 21 people participated in this test, including graduate students and researchers from several universities. The results of the listening tests are shown in Figure 5. The audio in *condition 1* is rated as the best quality while that in *condition 6* is rated the worst. The audio in *condition 3* gets the second-lowest rating, but assessors' opinions on the audio diverge most. Furthermore, by examining *conditions 1-3* or *4-6*, the more components of the F411E board are used, the lower the score for audio quality.

One reason for the different performance of Codec 2 on the F411E board and PC could be the floating-point precision. The F411E board only supports single-precision floating-point at the hardware level, while on PC, double-precision is supported. The difference in playback is apparent. One can hear a short periodic noise when listening to the output of the audio DAC on the F411E board. We primarily view this as an implementation and engineering issue, which we plan to address in our future work.

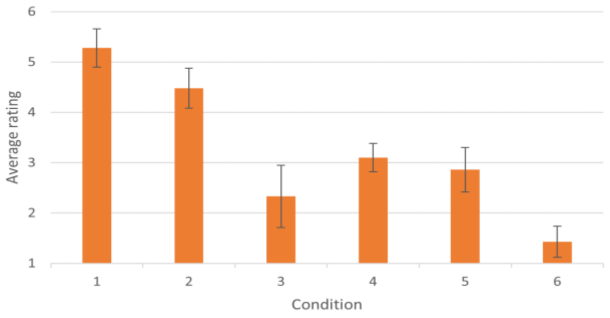


Fig. 5. Listening test results for audio quality under different conditions

2) Audio Quality under Packet Loss Situation: In real world, some packets may be lost during wireless transmission. It is natural to assume that a higher packet loss rate results in lower audio quality. We conducted the second quality test to verify this assumption. Packet loss is simulated by dropping part of the c2bits of an encoded audio clip with different loss rates and rebuilding audio from the c2bits. In semi-live or live audio applications, 14 bytes of payload are transmitted in each packet. The smallest unit to be dropped is 14 bytes. The loss rates tested are 10%, 20%, 30%, 40% and 50%. For each loss rate, c2bits are randomly dropped. In addition, a random seed value of 1000 is used to make the result reproducible.

This test is also delivered by questionnaire using Google Forms. The assessors first listen to reference audio, which is the audio from previous test *condition 1* because it got the highest quality rating. Then the assessors listen to five clips simulating lost cases in the packet and compare with the reference audio to give their opinions on the quality difference. There are five options for rating, numbers 1 to 5, with 1 indicating obviously worse than the reference and 5 indicating imperceptible compared to the reference.

The results are shown in Figure 6. With an increased loss rate, the corresponding average score goes lower. The 10% loss rate receives nearly a score of 5, while the 50% loss rate receives a uniformly lowest score of 1.

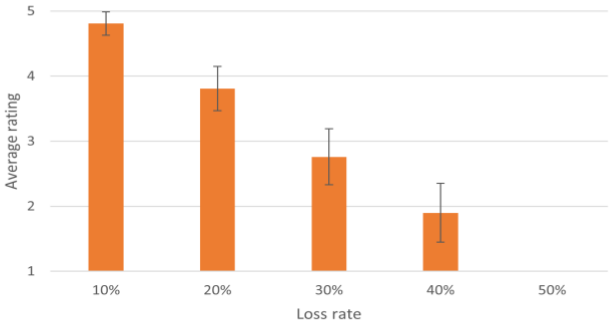


Fig. 6. Results of the listening test for the quality of lost packet audio

The results obtained confirmed the assumption that a high loss rate leads to low quality. In addition, a 10% loss rate has minimal impact on audio quality, where most assessors consider it imperceptible compared to reference audio. The reason could be that the 10% loss rate impact is too small to be recognised by most people since 10% less content does not change the essential information in the speech. In our view, these results indicate a bright future for ReSoNate, as packet loss rates above 10% are rare on well-engineered networks.

V. RELATED WORK

Nakamura et al. [12] added voice message functionality to a LoRa-based messaging system built by Cardenas et al. [2]. The core devices in the studies are called hubs, which have Wi-Fi and LoRa transceivers. The hubs provide connectivity to nearby devices via Wi-Fi and communicate with other hubs by LoRa. The system supports both broadcast and user-to-user modes. A user needs to register in the system to identify oneself so that messages destined for them can be received. In addition to sending a text message to users connected by hubs, the message can also be sent to an Internet message application, Telegram, via a gateway hub that links to the Internet.

For voice messages, the system first records input into an uncompressed WAV file. Then, FFmpeg [25] is used to convert the *wav file* to an mp3 file to reduce the message size, and the size is reduced to one-tenth of the original. The voice message finally is sent to a node of an MQTT system, and the subscriber to the corresponding MQTT topic will receive the message. The voice message experiment was done with transmission distances of 1m, 750m and 6000m. Performance is measured by successful transfer time (STT), the time from the first packet being sent to receiving the acknowledgement of the last packet. The result shows that distance had much less effect on transmission time than message size. A 100 Kbyte message containing 50 seconds of speech needs about seven minutes and a half, which might violate the duty cycle regulation for one hour.

Mekiker et al. [9] claimed that LoRa achieves point-to-point real-time voice communication in a proprietary implementation. They described a LoRa-based radio Beartooth along with the proposed Beartooth Relay Protocol aiming to support mobile application data and voice flow by LoRa. A Beartooth radio device has a Bluetooth transceiver to connect smartphones and a LoRa transceiver to connect other Beartooth devices. A multihop network can be established using multiple Beartooth radios. The source and destination devices are called nodes, while the devices in between are called relays. The Beartooth radio was responsible for the physical layer of LoRa, and an Android app on the smartphone handled the MAC layer. This approach is rather different to ReSoNate, which uses an unmodified version of the LoRaWAN stack running within the standard duty-cycle regulations. The protocol operates in cycles of two stages, negotiation and data exchange, and data is divided into two types: binary and voice. In the negotiation stage, a node first establishes a link and then sends requests to the relay. Next, the relay sends a transmission schedule to each requesting node, indicating which timeslot can be used by which node. Voice data has a higher priority in scheduling. In the data exchange state, nodes send data to the relay at assigned timeslots. In the throughput evaluation, the voice data rate was expected to reach 1.3 Kbps, which implied it could support Codec 2 1300 bps mode. Range evaluation shows that the Beartooth devices could maintain a connection of up to 30.4 km in line-of-sight conditions.

In [7], low bitrate audio compression with a bit rate of 64kbps was used, which is still higher than the LoRa 10 kbps data rate and cannot support live audio. Whereas, in Nakamura et al. [12], 50s audio with a size of 100 Kbyte is equivalent to the bit rate of 16 kbps, which is still not low enough to stream audio by LoRa. Furthermore, transmitting the 50s voice message at once might break the duty cycle regulation. Although live audio is achieved in Mekiker et al. [9] through a proprietary implementation, the paper presents no details about the live stream is realised. In addition, it follows the FCC regulation of LoRa frequency bands, where air time and maximum power restrictions are more relaxed than in the EU.

While we have not evaluated the range directly, we expect that our results would mirror prior work, as the achieved range is a characteristic of LoRa itself. Rather than the audio protocols that run on top of it.

VI. CONCLUSION AND FUTURE WORK

In this paper, we use the LoRa physical layer to explore the possibility of live audio transmission within the EU duty cycle regulations. ReSoNate demonstrates that Codec 2, in combination with a power-efficient wireless embedded platform can support real-time audio communication over the LoRa networks.

In terms of future work, the current study is limited by the amount and variety of speech sounds used in the audio evaluation. This could be improved by varying the duration of speech, speakers, or speed of speech. The promising features of ReSoNate pave way for future work in the following directions:

- The Codec 2 at 700 and 450 modes, requiring a lower data rate, could be used, although the audio quality might be worse. However, machine learning techniques can be explored to improve quality.
- Using a microphone with dual channels for the audio DAC input and supporting PCM format removes the PDM conversion requirements and reduces the processing load, thereby improving the time constraint issues impose by the buffer.
- Finally, we intend to invest significant engineering effort in improving the implementation of audio recording and playback on the embedded device, in order to address the shortcomings highlighted in Section IV.B.1.

ACKNOWLEDGEMENTS

The work presented in this paper is partially supported by the research fund, KU Leuven and the FWO-LOCUSTS project.

REFERENCES

[1] R. O. Andrade, and S. G. Yoo. A Comprehensive Study of the Use of LoRa in the Development of Smart Cities. In *Applied Sciences,* vol. 9, no. 22, pp. 4753, 2019.

[2] A. M. Cardenas, M. K. Nakamura Pinto, E. Pietrosemoli, M. Zennaro, M. Rain- one, and P. Manzoni. A low-cost and low-power messaging system based on the LoRa wireless technology. *Mobile networks and applications*, 25(3):961–968, 2020.

[3] W. Domski. SX1278. URL: https://github.com/wdomski/SX1278, last checked on 2022-05-30.

[4] 32F411EDISCOVERY - Discovery kit with STM32F411VE MCU - STMicroelectronics. URL: https://www.st.com/en/evaluation-tools/ 32f411ediscovery.html, last checked on 2022-06-07.

[5] European Commission, Directorate-General for Communications Networks, Content and Technology. Commission Implementing Decision (EU) 2017/1483 of 8 August 2017 amending Decision 2006/771/EC on harmonisation of the radio spectrum for use by short-range devices and repealing Decision 2006/804/EC (notified under document C(2017) 5464) (Text with EEA relevance).

[6] C. Gu, L. Jiang, and R. Tan. LoRa-based localization: Opportunities and challenges. *arXiv preprint arXiv:1812.11481*, 2018.

[7] R. Kirichek, V.-D. Pham, A. Kolechkin, M. Al-Bahri, and A. Paramonov. Transfer of multimedia data via LoRa. In *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*, pages 708–720. Springer, 2017.

[8] L. Li, J. Ren, and Q. Zhu. On the application of LoRa LPWAN technology in sailing monitoring system. In *2017 13th Annual Conference on Wireless On-demand Network Systems and Services (WONS)*, pages 77–80. IEEE, 2017.

[9] B. Mekiker, M. Wittie, J. Jones, and M. Monaghan. Beartooth relay protocol: Supporting real-time application streams over LoRa. *arXiv preprint arXiv:2008.00021*, 2020.

[10] K. Mekki, E. Bajic, F. Chaxel, and F. Meyer. A comparative study of LPWAN technologies for large-scale IoT deployment. *ICT express*, 5(1):1–7, 2019.

[11] Mitek. x893/codec2. URL: https://github.com/x893/codec2, last checked on 2022-05-30.

[12] K. Nakamura, P. Manzoni, M. Zennaro, J.-C. Cano, and C. T. Calafate. Adding voice messages to a low-cost long-range data messaging system. In *Proceedings of the 6th EAI International Conference on Smart Objects and Technologies for Social Good*, pages 42–47, 2020.

[13] C. Pham. Low-cost, low-power and long-range image sensor for visual surveillance. In *Proceedings of the 2nd Workshop on Experiences in the Design and Implementation of Smart Objects*, pages 35–40, 2016.

[14] A. H. Jebril, A. Sali, A. Ismail, and M. F. A. Rasid. Overcoming limitations of LoRa physical layer in image transmission. *Sensors*, 18(10):3257, 2018.

[15] J. Haxhibeqiri, E. De Poorter, I. Moerman, and J. Hoebeke. A survey of LoRaWAN for IoT: From technology to application. *Sensors*, 18(11):3995, 2018.

[16] D. Ilie-Ablachim, G. C. Pătru, I.-M. Florea, and D. Rosner. Monitoring device for culture substrate growth parameters for precision agriculture: Acronym: Monisen. In 2016 15th RoEduNet Conference: Networking in Education and Research, pages 1–7. IEEE, 2016.

[17] D. Sartori and D. Brunelli. A smart sensor for precision agriculture powered by microbial fuel cells. In *2016 IEEE Sensors Applications Symposium (SAS)*, pages 1–6. IEEE, 2016.

[18] C. Pham, A. Rahim, and P. Cousin. Low-cost, long-range open IoT for smarter rural African villages. In 2016 IEEE International Smart Cities Conference (ISC2), pages 1–6. IEEE, 2016.

[19] N. Poursafar, M. E. E. Alahi, and S. Mukhopadhyay. Long-range wireless technologies for IoT applications: A review. In *2017 Eleventh International Conference on Sensing Technology (ICST)*, pages 1–6. IEEE, 2017.

[20] P. A. Catherwood, D. Steele, M. Little, S. Mccomb, and J. McLaughlin. A community-based IoT personalized wireless healthcare solution trial. IEEE journal of translational engineering in health and medicine, 6:1-13, 2018.

[21] J. Petäjäjärvi, K. Mikhaylov, R. Yasmin, M. Hämäläinen, and J. Iinatti. Evaluation of LoRa LPWAN technology for indoor remote health and wellbeing monitoring. *International Journal of Wireless Information Networks*, 24(2):153–165, 2017.

[22] G.S. Ramachandran, F. Yang, P. Lawrence, S. Michiels, W. Joosen, D. Hughes. µPnP-WAN: Experiences with LoRa and its deployment in DR Congo, 2017 9th International Conference on Communication Systems and Networks, COMSNETS 2017, pages 63-70, IEEE, June 9, 2017.

[23] ARM Cortex-M4 32b MCU+FPU, 125 DMIPS, 512KB Flash, 128KB RAM, USB OTG FS, 11 TIMs, 1 ADC, 13 comm. interfaces. URL: https://www.st. com/resource/en/datasheet/stm32f411ve.pdf, last checked on 2022-06-07.

[24] A. H. Ahmar, E. Aras, T. D. Nguyen, S. Michiels, W. Joosen, D. Hughes. CRAM: Robust Medium Access Control for LPWAN using Cryptographic Frequency Hopping, DCOSS 2020, Distributed Computing in Sensor Systems: 16th IEEE International Conference, DCOSS 2020, 8 pages, Marina del Rey, CA, USA., May 25-27, 2020.

[25] FFmpeg. A complete, cross-platform solution to record, convert and stream audio and video. URL: https://ffmpeg.org/, last checked on 2022-09-30.

A Study of Data Augmentation for Chinese Character Data

Dong Bin Choi, Yunhee Kang, Myung-Ju Kang, Young B. Park*

***Abstract*— As Convolutional Neural Networks (CNNs) is making achievement in the field of optical character recognition (OCR), various languages are being generated as train data. However, in the case of languages that are not currently used, there are many difficulties in generating train data set. Traditional Chinese characters are also a language that is not currently used, but many documents remaining in East Asea are recorded in traditional Chinese characters, so recognition studies through OCR are being conducted. Data augmentation is used to generate lack of characters to make train data set. And some studies argue that conventional data augmentation is not sufficient. In this paper, we measured the difference in CNNs performance using only scaling and morphological deformation to generate train data set and find out whether such a claim is true. As a result, we were able to improve the performance of CNNs with 79.1% accuracy to 95.8%.**

***Index Terms*— Data Augmentation, Chinese Character Data, CNN.**

I. INTRODUCTION

Convolutional Neural Networks (CNNs) is the most studied deep learning architectures and has accomplished great achievements in the field of patter recognition including optical character recognition (OCR). However, to utilize CNNs, a lot of training data is required, but there are cases where the collected data is not enough. Languages that are not currently used, such as traditional Chinese character, even the colleting of data has limitations.

To overcome this problem data augmentation has been proposed. Although conventional data augmentation has been proven to be effective in improving CNN performance by applying it to multiple image data, there are studies that say it is not effective enough for languages such as traditional Chinese characters [1, 2].

This paper intends to investigate whether conventional data augmentation is effective or not for data such as traditional Chinese characters. Train data set was created by 72 types of traditional Chinese characters, and 5 letter for each type. Based on this train data set, scaling deformation were each or mixed to generate data, and we checked how the generated data affects the performance of a simple CNN with three convolution layers. In Section 2, we describe related research. In Section 3 we explain about the experiment, and in Section 4 presents the results. Finally, In Section 5 describes the future work.

II. RELATED WORK

2.1. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are special neural network architectures used for processing data with a grid like topology, such as one-dimensional time series or two-dimensional image data [3]. The origin idea of CNNs is from findings of Hubel and Wiesel's work on mammals primary visual cortex [4].

CNNs is using three architectures. local connectivity, weight sharing and pooling/subsampling. These helps to retain the spatial structure of data as well as ensuring some invariance towards affine transformations and distortions [5].

Using these CNNs Zhang et al. proposed a Chinese character recognition method and achieved a recognition rate of 97.3% using 720 training images and 60 evaluation images for each of 3,755 Chinese characters [1, 6].

2.2 Data Augmentation

For deep learning models to obtain satisfactory results they need to be fed a great deal of training data. Usually, more training data implies that the model can extract more relevant features and therefore become more robust.

In many cases, the datasets are not large or diverse enough and thus resulting in poor classification accuracy. A solution to this problem is to enlarge and diversify the data sets by augmenting them. This is known as data augmentation [3].

Conventional data augmentation is mainly four methods:

- affine transformation (rotation, translation, shearing & scaling),
- noise removal/injection (gaussian blur, gaussian noise & sharpening),
- morphological deformation (dilation & erosion),
- elastic distortion

And like Taihei Hayashi et al. or Xiwen Qu et al. there are some special methods for Chinese characters only [1, 2].

III. EXPERIMENT

Using 72 types and total number of 360 character base train data set were created. The letter was from the South Yang poetry book. The shape of the letters is shown in Figure 1.

Department of Computer Science, Dankook University, R.O.K. (email:dbchoi85@gmail.com)
Division of Computer Engineering, Baekseok University, R.O.K. (email:yhkang@bu.ac.kr)
Nectarsoft Co. Ltd., R.O.K. (email: kmjziro@nectarsoft.co.kr)
Department of Software, Dankook University, R.O.K. (email:ybpark@dankook.ac.kr)

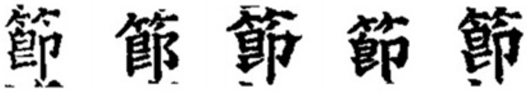


Fig. 1 Character from South Yang poetry book

The CNNs model structure used in the experiment is configured as shown in Figure 2 and consists of three convolutional layers. Figure 3 shows the results of learning with only 360 training data for reference points for comparison.

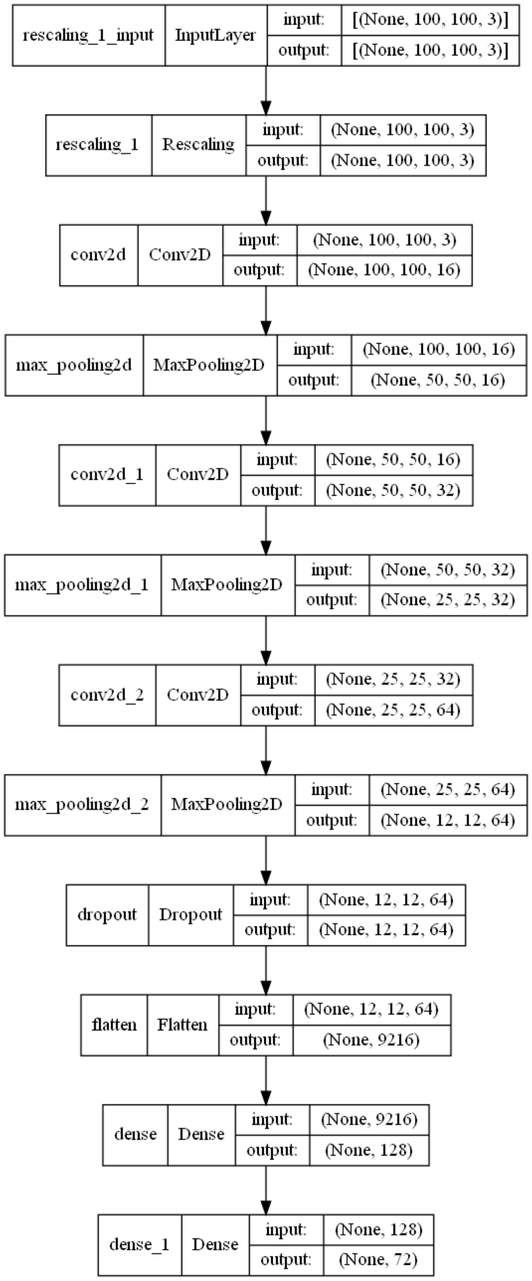

rescaling_1_input | InputLayer | input: [(None, 100, 100, 3)] | output: [(None, 100, 100, 3)]
rescaling_1 | Rescaling | input: (None, 100, 100, 3) | output: (None, 100, 100, 3)
conv2d | Conv2D | input: (None, 100, 100, 3) | output: (None, 100, 100, 16)
max_pooling2d | MaxPooling2D | input: (None, 100, 100, 16) | output: (None, 50, 50, 16)
conv2d_1 | Conv2D | input: (None, 50, 50, 16) | output: (None, 50, 50, 32)
max_pooling2d_1 | MaxPooling2D | input: (None, 50, 50, 32) | output: (None, 25, 25, 32)
conv2d_2 | Conv2D | input: (None, 25, 25, 32) | output: (None, 25, 25, 64)
max_pooling2d_2 | MaxPooling2D | input: (None, 25, 25, 64) | output: (None, 12, 12, 64)
dropout | Dropout | input: (None, 12, 12, 64) | output: (None, 12, 12, 64)
flatten | Flatten | input: (None, 12, 12, 64) | output: (None, 9216)
dense | Dense | input: (None, 9216) | output: (None, 128)
dense_1 | Dense | input: (None, 128) | output: (None, 72)

Fig. 2 Simple CNNs model structure

And for accurate comparison, two types of evaluation data were prepared. One is the letters of the South Yang Poems that are not used in the training data, and the other is the letters of ChongSwaeRok, which are completely different in shape, and consist of 72 letters, respectively. The data of the evaluation data is shown in Figure 4.

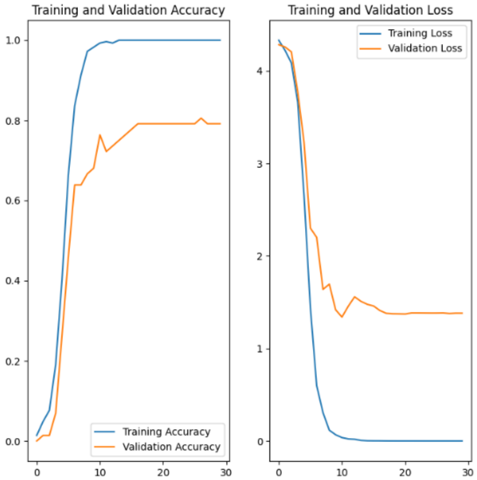


Fig. 3 Train result with base train data set

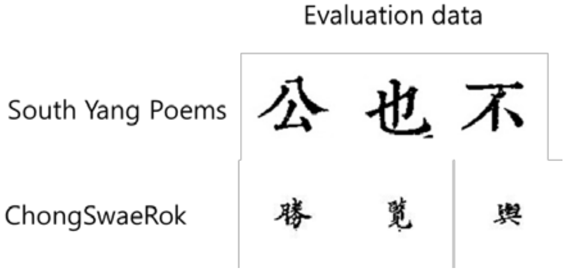


Fig. 4 Evaluation data

Table 1 shows the evaluation results of CNNs trained only with base train data set.

Table 1 Result of base train set

	South Yang	ChokSwaeRok
Right	57	1
Wrong	15	71

The first data augmentation method is scaling. Data augmentation was carried out by adjusting 7 steps for the scale of each word as Figure 5.

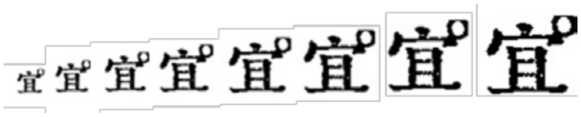


Fig. 5 Train data set generated by scaling

In this way, a total of 2880 train data set were generated, and the training result is shown in Figure 6.

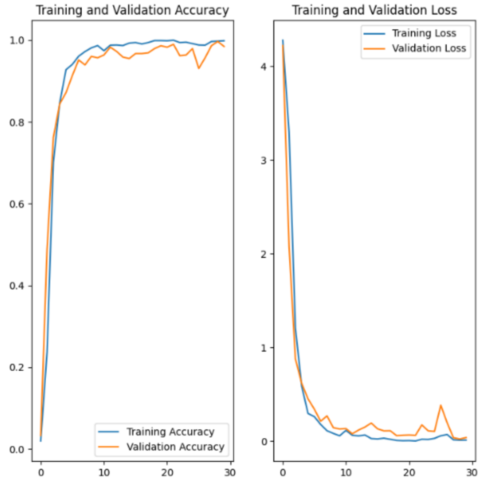


Fig. 6 Train result with scaling

And the evaluation results are shown in Table 2.

Table 2 Result of scaling

	South Yang	ChokSwaeRok
Right	63	35
Wrong	9	37

The second data augmentation method is morphological deformation. The form of the training data generated in this way is shown in Figure 7.

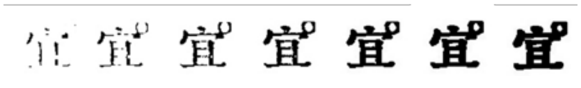


Fig. 7 Train data set generated by morphological deformation

A train data set consisting of a total of 2520 data was created, and the training results are shown in Figure 8.

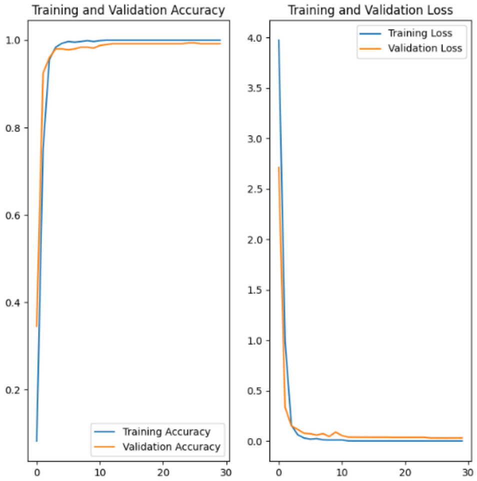


Fig. 8 Train result by morphological deformation

And the evaluation results are shown in Table 3 below.

Table 3 Result of morphological

	South Yang	ChokSwaeRok
Right	65	1
Wrong	7	71

In the last method, both methods were applied, and a train data set consisting of a total of 2,060 pieces of data was created. And the learning result is shown in Figure 9. The evaluation results are shown in Table 4.

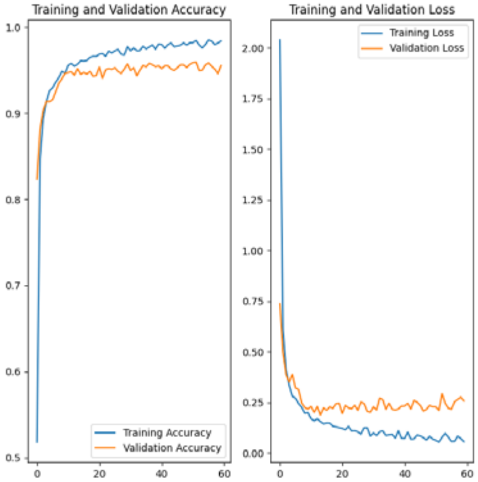


Fig. 9 Train result by complex

Table 4 Result of complex

	South Yang	ChokSwaeRok
Right	69	36
Wrong	3	26

IV. RESULT

Analyzing the experimental results, when each method was used alone, the data amplification method through scaling was the most effective, and the morphological deformation method was useful when evaluating similar data but was not effective when guessing data with a completely different shape.

The best result is to use both methods, but as a result, the amount of training data increases from 360 to 2060 in total.

There are more conventional data augmentation techniques not used in this paper, and the effect of each method is still unknown. Of course, data augmentation tailored to the shape of the learning data can be effective, but some effect can be expected by using the existing conventional data augmentation technique.

V. FUTURE WORK

It is necessary to understand the effects of the techniques not used in this paper on learning, and there is a need to analyze the differences between the techniques for Chinese characters and the existing techniques more closely.

Acknowledgement

This work was supported by Korea Evaluation Institution of Industrial Technology(KEIT) grant funded by the Korea government(MOTIE) (No.20022177, Open safety management service development based on AI collaboration technology between edge devices for high-risk site safety management)

References

[1] T. Hayashi, K. Gyohten, H. Ohki and T. Takami, "A Study of Data Augmentation for Handwritten Character Recognition using Deep Learning," 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2018, pp. 552-557, doi: 10.1109/ICFHR-2018.2018.00102.

[2] Xiwen Qu, Weiqiang Wang, Ke Lu, Jianshe Zhou, “Data augmentation and directional feature maps extraction for in-air handwritten Chinese character recognition based on convolutional neural network”, Pattern Recognition Letters, Volume 111, 2018, Pages 9-15, https://doi.org/10.1016/j.patrec.2018.04.001

[3] Bonnici, Elias, and Per Arn. "The impact of Data Augmentation on classification accuracy and training time in Handwritten Character Recognition." (2021).

[4] Hubel, David H and Wiesel, Torsten N. “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex”. In: The Journal of physiology 160.1 (1962), pp. 106–154.

[5] Lecun, Y. et al. “Gradient-based learning applied to document recognition”.In: Proceedings of the IEEE 86.11 (1998), pp. 2278–2324. DOI: 10.1109/ 5.726791

[6] Xu-Yao Zhang, Yoshua Bengio, Cheng-Lin Liu "Online and Offline Handwritten Chinese Character Recognition: A Comprehensive Study and New Benchmark", Pattern Recognition, vol. 61, no. 1 pp.348-360 (2017)

Smart Record and Transfer Videos to Different Targeted Audiences

Xinhang Xu, Yuxuan Zhao*, Yuechun Wang, Jie Zhang, Ka Lok Man

***Abstract*— In modern society, more emergencies and other unpredictable events can easily occur in daily life, leading to the heavy burden for human operators who must monitor these countless events on CCTVs with their own eyes. However, this situation can be changed with the dramatic development in object recognition enhanced by machine learning, which has the potential of releasing human operators' pressure and detecting possible emergencies automatically. This project has exactly been inspired to solve the problem with such kind of technology, which aims to develop an Android application processing videos that possibly contain emergencies such as fire, and send early alarms to users. The development of this project has been divided into three parts: designing the layout, training the machine learning model and realizing the functions on Android platform.**

***Index Terms*— Machine learning, Android development, image recognition, video processing.**

I. INTRODUCTION

In most situations, the CCTV videos are monitored by human operators; but with the increasing burdens, it could be truly tough for them to stay focused all day long. Since CCTV monitors are usually large in their numbers, operators have to switch frequently to watch stream videos of different locations. Thus it could be hard for them to spot possible emergencies immediately. To deal with these problems, this project, which is an Android application, is designed and implemented to detect possible fires in videos uploaded to mobile devices. It will report the detected danger to users by automatically analyzing videos in the background. To avoid confusion, this project has been implemented to deliver fire warning messages only to those professionals who deal with fire emergencies. Last but not least, if the warning is a fake one and users want to dismiss this warning, they have to make double check to ensure that this fake warning is caused by the wrong detection.

II. LITERATURE REVIEW

In this project, Android OS has been selected to be the running platform mainly because of two reasons: first and foremost, Android OS has taken up nearly three fourth of the total mobile device market in recent years [1], thus building such an application on Android platform has the potential of benefiting much more users; secondly, Android development has a large variety of built tools and API to invoke for conducting automatic testing on the current code, avoiding possible mistakes at the initial stages [2]. However, since this project is focused on implementing emergency detections, it would probably acquire and process people's privacy and other data which should be accessed by special groups of people such as policemen and firefighters, while the Android platform still faces large threats of leaking its security permissions to unwanted users [3].

This project is designed for processing possible fires in videos. However, the typical video processing procedures contain acquiring the sampling images, or "frames" in the videos at first, then analyzing the frames [4]. This processing method can simplify the procedures from analyzing continuous videos to discrete pixels' behaviours only, reducing time and energy cost while ensuring a relatively quick responding time, which is truly valuable for mobile devices which do not have the same computing resources as PC and workstations do. For the detection session, the Yolov5 model has been utilized in this project owing to its one-stage characteristic, which is quite suitable for real-time detection with its fast speed [5]. Among the Yolov5 model family. Yolov5s has the smallest volume with only 27MB, and can easily be deployed on mobile devices [6].

III. METHODOLOGY

The project can be realized by methodology separated into three parts: first is the image capturing function, which acquires the screenshot images of the videos uploaded to this app at fixed time intervals. This step is for achieving higher efficiency in detection since the app only needs to analyze single images rather than the whole video. Second is the machine learning model, which is trained with proper datasets and capable of detecting fire in most videos uploaded, except for those with too low resolutions. The last part is the Android software, which integrates all the functions together.

The flowchart of this whole methodology is shown below:

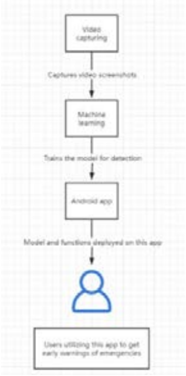

Video capturing
Captures video screenshots
Machine learning
Trains the model for detection
Android app
Model and functions deployed on this app
Users utilizing this app to get early warnings of emergencies

Fig. 1 Flowchart of the project's applied methodology

A. Android Application

Because of its feature of allowing developers to flexibly achieve various functions [7], the Android platform has been chosen to realize the objectives of this project. To support as many devices as this project can, the SDK 24 (Android Nougat) has been used as the lowest supported version. Although it only supports 73.7% existing Android devices, which is relatively lower compared to other frequently selected versions such as SDK 21 (Android Lollipop), which supports 94.1% devices [8], choosing this version is still of necessity since it has better support for deployment of the Yolo algorithm series, and fewer bugs can be spotted during the procedure of development [9].

B. Machine Learning Model

In this project, Yolov5 has finally been selected to train the detection model. Yolo is a one-stage algorithm model, and the "yolo" here refers to "you only look once". This is in contrast with the traditional R-CNN series algorithms, which have the "two-stage" network structure. This structure has allowed R-CNN to detect items with higher accuracy while its running speed cannot satisfy the requirement of real-time detection [10].

Normally, Yolo processes images with three steps: it firstly resizes the image for detection into a 448×448 size one (which has been reduced to 416×416 in yolov2), then it activates the convolution neural network, and lastly, it determines whether the expected item exists in this image according to the trained model's thresholds [11].

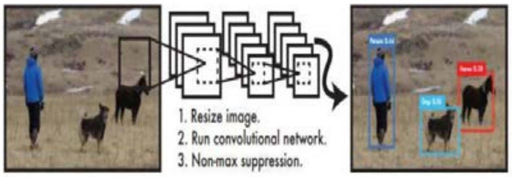


Fig. 2 Steps of yolo detection [11]

For the image input, Yolo firstly divides it into S×S cells. If an object's center locates in one of these cells, then this cell will be responsible for detecting this object and determining the bounding box, which can round up the detected object. The cell will also predict the confidence score, which is composed of two elements: one is the possibility of target object existing in the current cell, the other is how accurate the position of the bounding box is. Adding the confidence score to the position information presented by four values x, y, w, h (the x-coordinate, y-coordinate, width and height), each bounding box will predict five values in total. Each cell should also predict class information, which should have been stated during the training procedure. To sum up, we set each cell to predict B bounding box, and set the number of classes to be C. Then for S×S cells, the final tensor output size will be S×S×(B×5+C) (For detecting more classes at a time, in yolov2, the class number has been added in each cell) [6].

From Yolov2, determining the position of the bounding box no longer depends on the four values x, y, w and h only. Since Yolov1 does not have a recommendation area just as R-CNN does, Yolov2 improved on this part by adding representative prior anchors to make the network converges more easily [13]. For the working network, Yolo has been improving during its five editions. For Yolov1, a typical one-stage convolutional neural network was built, with the input of 448×448×3 images at the input side, followed by several convolution layers and maximum pooling to extract the abstract characteristics of images in the middle layer, along with two full connected layers to predict target location and class probabilities. The 7×7×30 prediction output comes at last [11].

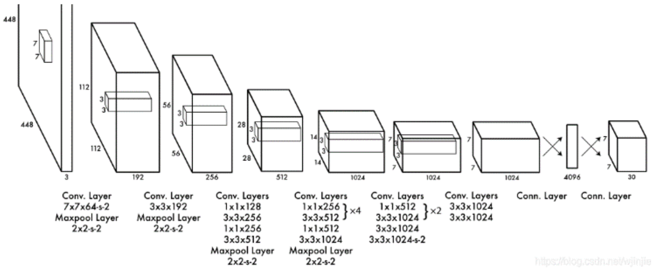
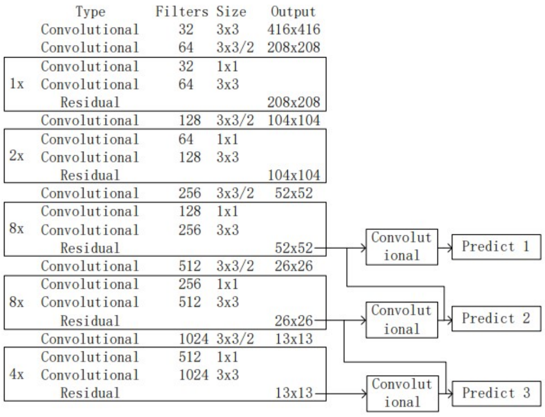


Fig. 3 Network structure of yolov1 [11]

In yolov2 and yolov3, however, Darknet has been introduced to conduct the feature extraction function. Compared with yolov1, after each layer of convolution, batch normalization has been added to do pre-processing to improve the effectiveness of the system. Moreover, 1×1 convolution has been set between those 3×3 ones to compress the features to save more space. Since Yolo has the ability of detecting items even with low accuracy training sets, this could strengthen its advantage of doing real-time detection [11][15].



	Type	Filters	Size	Output
	Convolutional	32	3x3	416x416
	Convolutional	64	3x3/2	208x208
1x	Convolutional	32	1x1	
	Convolutional	64	3x3	
	Residual			208x208
	Convolutional	128	3x3/2	104x104
2x	Convolutional	64	1x1	
	Convolutional	128	3x3	
	Residual			104x104
	Convolutional	256	3x3/2	52x52
8x	Convolutional	128	1x1	
	Convolutional	256	3x3	
	Residual			52x52
	Convolutional	512	3x3/2	26x26
8x	Convolutional	256	1x1	
	Convolutional	512	3x3	
	Residual			26x26
	Convolutional	1024	3x3/2	13x13
4x	Convolutional	512	1x1	
	Convolutional	1024	3x3	
	Residual			13x13

Fig. 4 Network structure of yolov3 [15]

For Yolov4 and Yolov5, with the latter one of which has been applied in this project, they have mainly improved in several aspects.

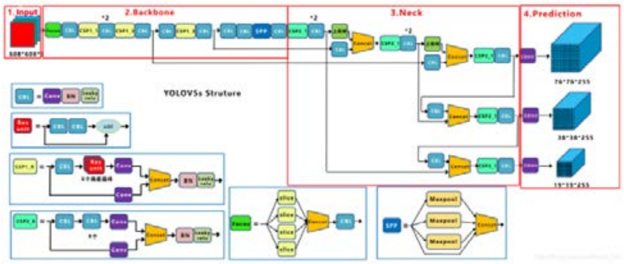


Fig. 5 Network structure of yolov5 [16]

First and foremost, the input images are enhanced by Mosaic data enhancement. This method applies random resizing, cropping and arranging patterns of different images, making the detection target's information more abundant for the system to predict. It can also increase the detection accuracy of small objects. At the backbone part of yolov5, it mainly applies Focus and CSP structures. Focus structure is important for its slicing operation. In this project, the light-level yolov5s has been utilized, which input 608×608×3 images into the Focus, then slice the images into 304×304×12 featured images and go through a 32-level convolution to become 304×304×32 featured images [16]. The neck part of yolov5 utilizes PANet. In this model, three frameworks have been proposed: first of all is the bottom-up path augmentation. Since neurons at higher levels have strong responses to the entire object, and those at lower levels can be activated by partial textures more easily, a top-down segment has been added to FPN network to deliver semantically strong features. The PANet adds a bottom-up path to improve the classification ability of the whole feature layer by delivering information at lower layers [17]. The second one is adaptive feature pooling, which is a structure mapping each proposal to a different feature level, then executing the ROIAlign. This procedure is aimed at providing proposals with useful semantic information to make predictions [17][18].

The last one applied is the fully-connected fusion. Since for FCN layer, it predicts each location's information based on different parameters, it has the capability of adapting to different locations, while its predictions are made according to the overall information of the entire proposal, which is useful when differentiating entities and different components of the same object. These two advantages will be combined by this full-connected fusion by predicting the binary pixel-wise mask for each class in order to decouple the mask and its class [17][18].

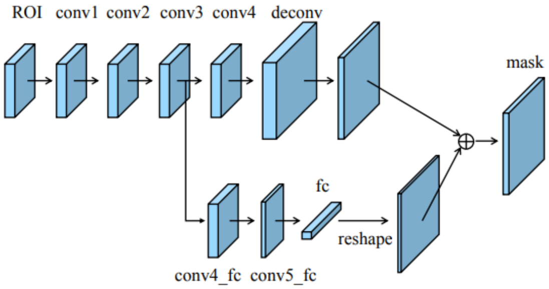


Fig. 6 The mask prediction path with fully-connected fusion [17]

IV. EXPERIMENTAL RESULTS

In this project, to evaluate whether this Android application is implemented appropriately, three steps have been applied: First of all, whether the video screenshots are captured at fixed intervals is tested on the Android platform with a mobile device. As mentioned before, this step releases the burden of the detection system by only detecting images rather than the whole videos. Secondly, the detection system which applies the yolov5 model will be evaluated to determine whether this system is capable of locating fire with enough accuracy. This step is realized by observing its theoretical data during its training procedure and using test videos with or without fires, which are selected randomly from the Internet and are not in the original training set. Lastly, the whole app will be encapsulated to evaluate its working precision and efficiency in real working situations. In this procedure, various operations which can be conducted by real users will be tested on this app to see whether there is any bug or disturbance, considering the fact that image recognition tends to require large amounts of computing resources, and has the possibility of causing some lagging from time to time. The testing video of this app is stored in this link: https://box.xjtlu.edu.cn/smart-link/0d280f60-e481-47a2-b439-147682ab0dd3/

A. *Screenshot Capture*

Since multiple videos are processed at the same time in this app, it has to be ensured that such multi-tasking can be conducted smoothly. Therefore, the first function to be checked is screenshot capturing, which takes a screenshot for each video every two seconds for the system to do further detection. Videos with different resolutions have been downloaded from the Internet to test the capturing. The results showed that for different videos which are processed at the same time, the app did not occur any lagging. Moreover, the screenshots have been captured at expected intervals after comparing them with the initial videos' time stamps.

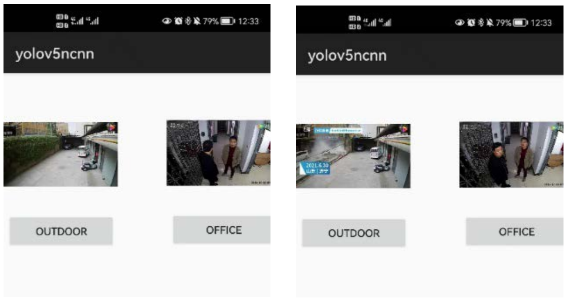


Fig. 7 Testing results of video capturing (with two seconds' interval)

B. *Fire Detection*

The detection of fire is the main focus of this project, whose results are vital for the safety issues to be dealt with this app. This app is designed to detect as many fires as possible instead of easily ignoring some possible fire warnings, that is to say, false negatives are much less acceptable than false positives considering people's safety.

In order to evaluate the results of detection, two methods have been applied:

First is to observe the training results on TensorBoard, which is a visualization interface for model training [19]. Since in yolov5, it trains the model iteratively with all the images and labels in the training set, the precision and recall are suitable indicators for the outcomes of the trained model. However, during the first training procedure, the precision and recall presented a decreasing tendency with the growth of the training loop number.

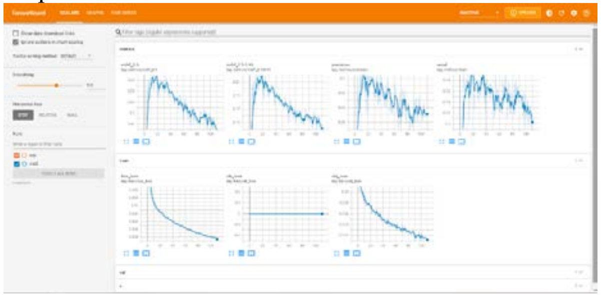


Fig. 8 First training results of the model

After analyzing possible reasons, we discovered that this might be caused by too complex labelling. Since I used polygons rather than rectangles to locate the fire on training images using Anaconda Navigator, which could cause trouble for the model to study from the data. Moreover, the surrounding interruptions around the fire have been simply eliminated so that the model may face some troubles when it is provided with a real image to detect.

We fixed this problem by labelling again, in which a new training set with fire images under more circumstances has been applied, and I switched to rectangles for labelling the training images. The detection precision has risen up to 95.8% after 300 epochs of training, which has been tested to be enough when detecting fire in daily videos.

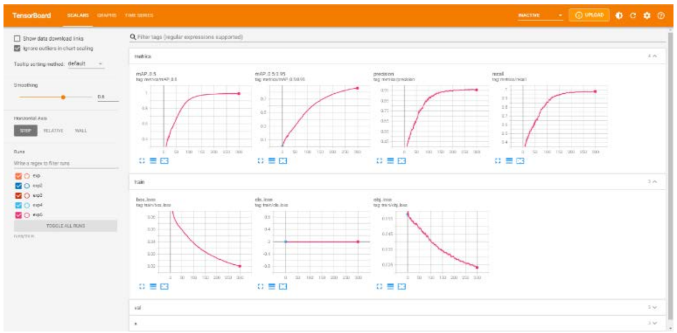


Fig. 9 Final training results of the model

The images for the training batches are also displayed as follows:

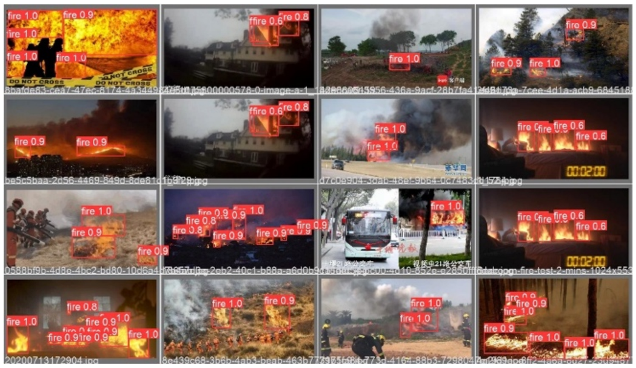


Fig. 10 Training batches of the fire

The second method applied is testing the trained model with a set of videos, which are randomly gathered from the Internet and have not appeared in the original training set. This testing procedure is aimed at determining two things: first is the detection accuracy in different videos. Since the resolutions of testing videos can vary a lot, the fires appeared in each of them can also show different shapes or features. This step is to ensure that all these fires can be detected correctly. The second is the mis-detection rate. Many videos without fire will also be included in the training set, with some of them containing objects similar to fires, such as sunshine on the wall. This step focuses on figuring out whether the rate of false positive is under a reasonable level.

From this step, it can be concluded that the trained model has reached a relatively high detection accuracy, given the fact that it has successfully detected all the fires from each testing video, with less than 0.5 seconds delay at most from the initialization of fires. For those objects showing some shared features with fires, this model has also proven itself to be error-resistant, considering that it has not been fooled by these similar objects under 95% circumstances, and even in the other 5% circumstances, the prediction values are also at a lower level compared with the established threshold, which is 0.70. Some of the testing results are shown in the following screenshots:

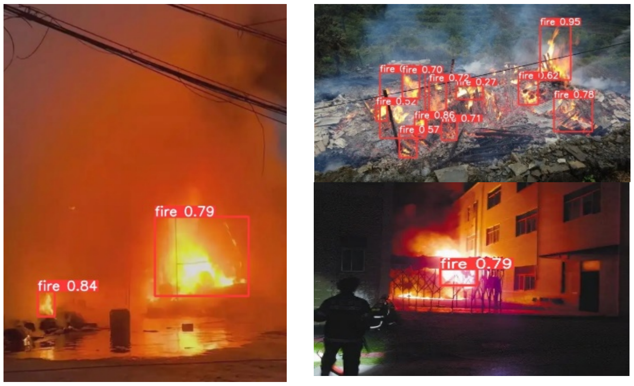


Fig. 11 Testing results in videos with fires

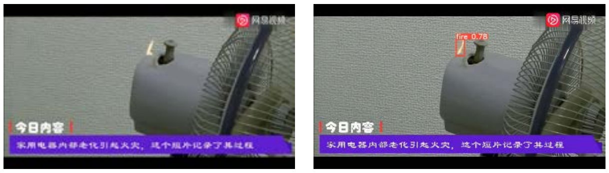
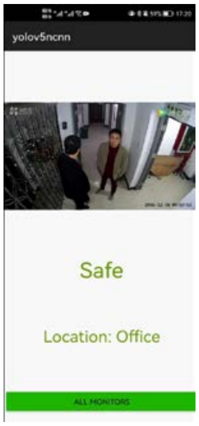


Fig. 12 Maximum delay of 0.5 second after the fire starts (Left image shows the beginning of the fire, right image shows the detection after 0.5 second)

C. *Application Test*

After the two components of capturing screenshots and the training model have been completed, the final evaluation will be on the whole Android application, which has encapsulated the two components. Multiple operations will be conducted to see whether this app can support the main functions well, and provide users with warnings as early as possible when it detects a possible fire.

Videos are captured in the initial page every two seconds. If no fire is detected in this video, this page will display a "Safe" text, along with the location where this video is recorded. There is a button called "All Monitors" at the bottom of this page. By clicking on this button, users can enter another page, which contains all the video screenshots being updated in real time at fixed intervals.


yolov5ncnn

Safe

Location: Office

Fig. 13 The initial page of this app

By clicking on each button in the screenshots, users can enter the corresponding individual page, in which they can observe the enlarged images captured from the videos. Each time when users switch from one page to another, the detecting procedure will be refreshed in order to save computing power and avoid delay. Switching between these pages has been tested to see whether this refreshment has been successful. The results have demonstrated that by properly refreshing the function, lagging in this application can be effectively prevented. When a fire is detected, the whole application will immediately jump to the corresponding individual page, on which the fire in the screenshot will be labelled out.

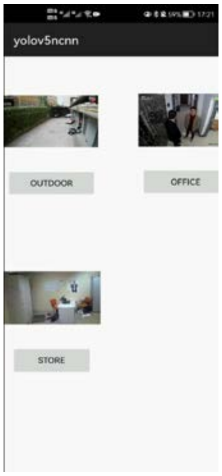
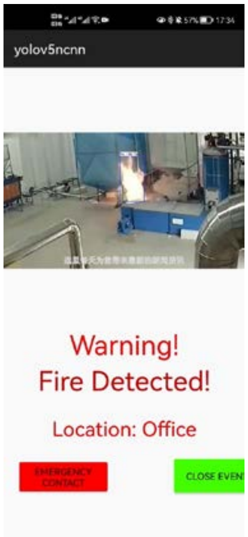

yolov5ncnn

Fig. 14 Page displaying the screenshots of all the videos


yolov5ncnn

Warning!
Fire Detected!

Location: Office

Fig. 15 Warning displayed when a fire is detected

The last function to be tested is the double-check. If the warning is false after being manually checked by human operators, and they want to dismiss it, they can click on the dismiss button. However, since emergencies can do great damage to people's safety and properties, dismissing the warning should be carefully done. Thus, this app provides a double-check function for users: when they click the dismiss button, a pop-up window will appear to ask whether they truly want to dismiss it.

V. CONCLUSION AND FUTURE WORK

A. *Conclusion*

In this project, the main functions of monitoring videos and detecting fire have been realized without bugs. Through the testing procedure, it has been evaluated that the detection precision is relatively high in correspondence to what the model has indicated during its training process. However, when the application encounters some interruptions, such as sunshine, it may cause misjudgments for a short while. Moreover, the application is not quite capable of detecting the initial fire in low-resolution videos.

B. *Future Work*

More types of emergencies can be trained in the model to help more operators deal with different types of emergencies. Datasets such as UCF-101 can be utilized to capture and detect human motions in other unpredictable events such as riots. For the non-functional part, the present interface is relatively simple, without a user login to make identifications. Since this project aims at helping monitor operators in special fields, the data and message should be more secured in this way. Also, the practical needs and using habits of these operators should be interviewed and implemented in the future. The proposed work will be potentially tested in the trusted environment in the future.

VI. ACKNOWLEDGEMENT

This work is partially supported by the Xi'an Jiaotong-Liverpool University (XJTLU) AI University Research Centre, Jiangsu (Provincial) Data Science and Cognitive Computational Engineering Research Centre at XJTLU; and research funding: XJTLU-REF-21-01-002.

REFERENCES

[1] Y. Yao, W. Jiang, Y. Wang, P. Song, and B. Wang, "Non-functional requirements analysis based on application reviews in the android app market," Information Resources Management Journal, vol. 35, no. 2, pp. 1–17, 2022.

[2] F. N. Musthafa, S. Mansur, and A. Wibawanto, "Automated software testing on mobile applications: A review with special focus on Android platform," 2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer), 2020.

[3] G. Shrivastava, P. Kumar, D. Gupta, and J. J. Rodrigues, "Privacy issues of Android
Application Permissions: A literature review," Transactions on Emerging Telecommunications Technologies, vol. 31, no. 12, 2019.

[4] V. Sharma, M. Gupta, A. Kumar, and D. Mishra, "Video processing using Deep Learning Techniques: A Systematic Literature Review," IEEE Access, vol. 9, pp. 139489–139507, 2021.

[5] J. Miao, G. Zhao, Y. Gao, and Y. Wen, "Fire detection algorithm based on improved Yolov5," 2021 International Conference on Control, Automation and Information Sciences (ICCAIS), 2021.

[6] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of Yolo algorithm developments," Procedia Computer Science, vol. 199, pp. 1066–1073, 2022.

[7] A. Almisreb, H. Hadžo Mulalić, N. Mučibabić, and R. Numanović, "A review on mobile operating systems and application development platforms," Sustainable Engineering and Innovation, vol. 1, no. 1, pp. 49–56, 2019.

[8] Google, Android SDK version properties [Online]. Available: https://developer.android.com/ndk/guides/sdk-versions.

[9] Ultralytics, "Ultralytics/yolov5: Yolov5 in PyTorch &gt; ONNX &gt; CoreML &gt; TFLite," GitHub. [Online]. Available: https://github.com/ultralytics/yolov5.

[10] S. D. Achar, C. Shankar Singh, C. S. Sumanth Rao, K. Pavana Narayana, and A. Dasare, "Indian currency recognition system using CNN and comparison with yolov5," 2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC), 2021.

[11] Y. Zhang, X. Li, F. Wang, B. Wei, and L. Li, "A comprehensive review of one-stage networks for object detection," 2021 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), 2021.

[12] J. Redmon, S. Divvala, R. B. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection." 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 779-788, 2016.

[13] P. Garg, D. R. Chowdhury, and V. N. More, "Traffic sign recognition and classification using yolov2, faster RCNN and SSD," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019.

[14] H. Zhang, L. Qin, J. Li, Y. Guo, Y. Zhou, J. Zhang, and Z. Xu, "Real-time detection method for small traffic signs based on yolov3," IEEE Access, vol. 8, pp. 64145–64156, 2020.

[15] F. Lin, X. Zheng, and Q. Wu, "Small object detection in aerial view based on improved Yolov3 Neural Network," 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), 2020.

[16] "Yolov5 Network Structure Studying," yolov5_Network Structure Studying. [Online]. Available: https://blog.csdn.net/Sept_Oct/article/details/115863842. [Accessed: 09-May-2022].

[17] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path Aggregation Network for instance segmentation," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.

[18] X. Zhang, H. Fan, H. J. Zhu, X. Huang, T. Wu, and H. Zhou, "Improvement of YOLOV5 model based on the structure of Multiscale Domain Adaptive Network for crowdscape," 2021 IEEE 7th International Conference on Cloud Computing and Intelligent Systems (CCIS), 2021.

[19] J. Yan, T. Liu, X. Ye, Q. Jing, and Y. Dai, "Rotating Machinery Fault diagnosis based on a novel lightweight convolutional neural network," PLOS ONE, vol. 16, no. 8, 2

EmotionFooler: An Effective and Precise Textual Adversarial Attack Method with Part of Speech and Similarity Score Checking

Fan Yang, Erick Purwanto*, Ka Lok Man

The accuracy rate of natural language processing models is the main pursuit of researchers. Although popular models can achieve a high accuracy rate, they are easy to be attacked when inputting wrong or misleading information, which indicates low robustness. Adversarial attacking is a useful method to increase the robustness of a model. Currently, most of the adversarial generation models create adversarial samples by substituting some words with their synonyms [1]. TextFooler is one of these models and it is achieved by two main mechanisms, "Word Importance Ranking" and "Word Transformer" [2]. However, it is easy to be attacked by adversarial samples because of the defects of "Word Transformer" algorithm. We provide an improved model mainly focuses on part of speech and similarity score checking, which is called EmotionFooler. After analyzing and evaluating the results, our model improves the quality of generated samples. EmotionFooler shows better results in two tasks of text classification and natural language inference and focuses on attacking BERT [3], WordLSTM [4] and WordCNN [5] with datasets MR, IMDB and Yelp. The attacking results are more natural and comprehensible by applying algorithms of speech evaluator, similarity score limiter and stop words list.

Index Terms— **Deep learning, Adversarial examples, Natural language processing, Textual attack.**

I. INTRODUCTION

A. Motivation, Aims and Objective

With the increasingly booming applications about natural language processing (NLP) scattered around the artificial intelligence and computer industry, the security of NLP models has aroused great concern. It is obvious that the error rate of NLP models greatly increases when inputting some special data samples whose uniqueness is undetectable to human beings [6]. These input samples are called adversarial samples because they have the ability to mislead the model and cause it to produce wrong output to increase robustness, as in Fig. 1. Rather than attacking models, the ultimate goal is to increase the robustness by training models with these adversarial samples.

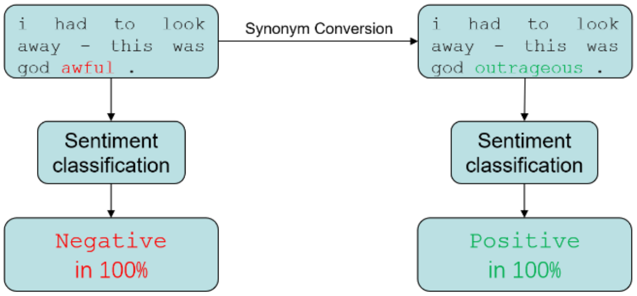


Fig. 1. Our Adversarial attack on BERT.

However, generating such samples is difficult because semantics, similarity and rationality about words or sentences should be taken into well consideration [7]. For instance, the mainstream view among researchers is replacing candidate words with their synonyms in a sentence. It is regarded as the most efficient way to produce adversarial sentences. Nevertheless, in human languages, one word can be changed not only to its synonyms, but sometimes can also be replaced with candidate words of different part-of-speech (POS), phrases or a word that has no relationship with the original one. Therefore, a well-designed adversarial model is necessary and also required to satisfy different demands.

In order to increase robustness, three main improvements will be made. First, the most common error in generated results is that the original model replaces words into synonyms with different part of speech. Therefore, we will pay attention to the syntax and semantic errors of replaced words in generated samples, which aiming to find the defects of the original model. After that, the algorithm of part-of-speech evaluator is introduced to efficiently analyze and improve the generated sentences from samples. Second, the stop words list will be checked and ameliorated because of the errors of wrong replacement. Third, compared with original generated samples, a similarity limiter will be used to limit the similarity score and select synonyms with high quality. Finally, compared with original generated samples, new samples will show higher quality and more natural sentence in result tables.

B. Literature Review

In order to test whether the accuracy rate of a model is falsely high, textual adversarial attacking has increasingly developed in the area of natural language processing. With advanced well-designed mechanisms and algorithms, the quality of adversarial

All authors are with the School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, R.O.C. (email: Fan.Yang18@student.xjtlu.edu.cn, {Erick.Purwanto, Ka.Man}@xjtlu.edu.cn).

samples generated by models is getting much higher and these samples are able to cause a large number of errors in NLP models with low robustness, resulting in a sharp decline in the accuracy rate.

In 2015, J. Goodfellow denied the previous view that misclassify arbitrary examples were regarded as the cause of model errors, but proposed that using adversarial examples to train models will reduce these errors [8]. This is a novel research method and the author found it has the ability to increase the robustness of a model. Based on it, HotFlip [9] was created to generate write-box adversarial samples for a classifier model. It indicates the great potential of adversarial samples in attacking models and improves models' robustness. Moreover, high successful attacking rate (100% on IMDB dataset) was achieved in the generation model of TextBugger [10] because of the combination of both white-box and black-box settings. Furthermore, after encountering the bottleneck in improving the accuracy rate, other researchers from Europe presented a method of inserting single adversarial word into a sentence [11]. Word-level Attacking [12] proposed a novel method and solved the issue of inappropriate search space reduction and the low efficiency of optimization.

The above models produce adversarial samples by relatively fixed algorithms and replace the word with its synonym. Currently, some new adversarial attacking methods of natural language processing have been created. For instance, TextFooler [2] has two main mechanisms to generate adversarial samples (Word Importance Ranking and Word Transformer) and it successfully attacked popular NLP models like BERT [3], WordLSTM [4] and WordCNN [5]. Although it also concentrates on changing words into their synonyms, the view of words with different part-of-speech can be instituted is presented in this paper. Nevertheless, some of the samples it generates are not easy to understand and strange to human' languages. To improve this, CLARE [12] alleviated the issues of unnatural and ungrammatical samples by three main methods of Replace, Insert and Merge. Moreover, the model BAE [13] increased the quality of results in aspect of syntax and semantic.

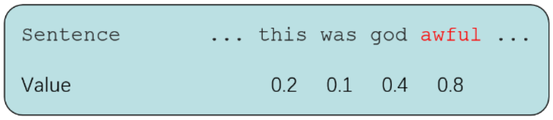
II. METHODOLOGY AND RESULTS

A. Methodology

1) Methodology of the original model: TextFooler contains several important parts. One is importance ranking, which is used to select the most K important words in a sentence.

$$
\begin{aligned}
I_{wi} &= F_Y(X) - F_Y(X_{wi}),\ if\ F(X) = F(X_{(}w_i)) = Y \\
&= (F_Y(X) - F_Y(X_{wi})) + (F_Y(X_{wi}) - F_Y(X)), \\
&\quad if\ F(X) = Y, F(X_{wi}) = Y, Y! = Y
\end{aligned} \tag{1}
$$

The main mechanism of importance ranking in Fig. 2 is that we mask one word in a sentence to make a new sentence. Then, the prediction score will be calculated after putting two sentences into the original model. After that, subtracting the two prediction score. Thus, the ranking score is exactly the difference of their prediction score, as shown in Function 2.



Sentence	...	this	was	god	awful	...
Value		0.2	0.1	0.4	0.8	

Fig. 2. Word Importance Ranking Sample.

$$
\begin{aligned}
V_{Candidate:awful} = {} & F_{Model}(Sentence) - \\
& F_{Model}(Sentence_{\backslash awful})
\end{aligned} \tag{2}
$$

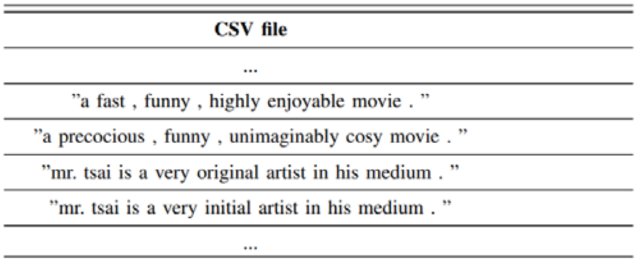
After that, using the cosine similarity to find the synonyms of the important words. In the process of this, part-of-speech checking and semantic scoring method will be applied to delete some unsuitable words from the important words list.

In our experiment, the dataset CoLA [14] will be used to test the performance of BERT, ALBERT and RoBERTa. The results will be presented as a table with four parameters, original accuracy, adversarial accuracy, adversarial changed rate and number of queries. This is regarded as the original results. Moreover, SST-2 (BERT-Base-Uncased) will be chosen as a pretrained model to generate 100 attacking samples, which is used to test our new algorithms. It is convenient because the size of the dataset is proper and the sentence complexity is also within the requirements of this experiment. After that, the similar mechanisms such as automatic evaluation and human evaluation will also be used to evaluate the quality of generated samples. All the pretrained models and datasets are downloaded from HuggingFace [15].

2) Using algorithm to locate errors: After implementing the model, large numbers of samples are generated from different pretrained language models and the corresponding datasets. On the whole, these outputs showed high attacking success rates and they have demonstrated the remarkable achievements in generating offensive text. However, although the generated examples perform well generally, when we carefully observe each generated sentences, there do exist some errors which has a negative impact on the quality of the whole samples. Therefore, an algorithm will be used to detect errors.

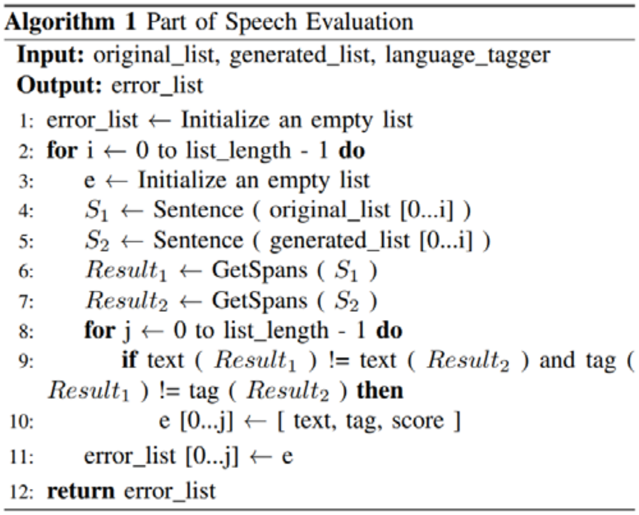
For the algorithm, all the generated samples will be uniformly formatted and neatly arranged in the CSV file as Table I. In the file, odd lines represent the original sentences, while newly generated sentence through the model are in even lines.

After that, the aim is to read the CSV file and process these sentences by using the Algorithm 1 Part of Speech Evaluation. Our method will call a natural language processing toolbox called "flair", which helps to analyze the components of the whole sentence. We will use the "0.10" version of this tool, which has good compatibility and efficiency. Inside the code, the functions "Sentence" and "GetSpan" are used and they will return the part of speech of each word and the usage rate of part of speech of the word. The next step is to iterate the data and find the words have been replaced by the model with different part of speech with the original ones. It is a way of vital importance to judge whether there exist grammatical and semantical errors.



CSV file
...
”a fast , funny , highly enjoyable movie . ”
”a precocious , funny , unimaginably cosy movie . ”
”mr. tsai is a very original artist in his medium . ”
”mr. tsai is a very initial artist in his medium . ”
...

TABLE I: Generated samples in CSV file

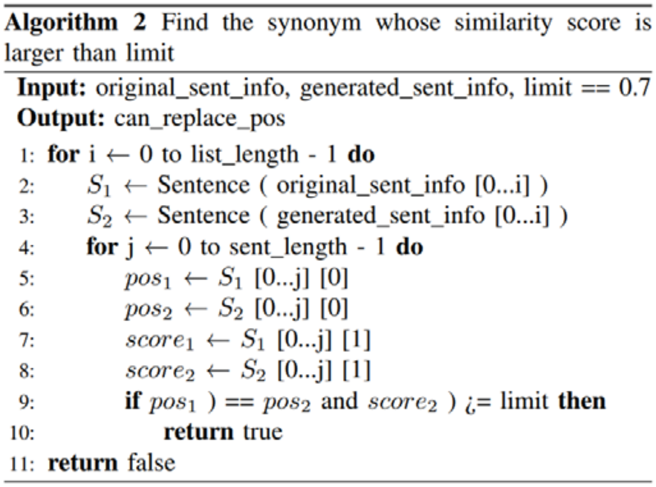


```
Algorithm 1 Part of Speech Evaluation
Input: original_list, generated_list, language_tagger
Output: error_list
error_list ← Initialize an empty list
for i ← 0 to list_length - 1 do
    e ← Initialize an empty list
    $S_1$ ← Sentence ( original_list [0...i] )
    $S_2$ ← Sentence ( generated_list [0...i] )
    $Result_1$ ← GetSpans ( $S_1$ )
    $Result_2$ ← GetSpans ( $S_2$ )
    for j ← 0 to list_length - 1 do
        if text ( $Result_1$ ) != text ( $Result_2$ ) and tag ( $Result_1$ ) != tag ( $Result_2$ ) then
            e [0...j] ← [ text, tag, score ]
    error_list [0...j] ← e
return error_list
```

In the next step, human evaluation is applied to evaluate the generated samples to find out some special cases with errors. Some parameters such as replaced rate, error rate, number of replaced words and replaced POS tags will be listed as a table for analyzing the model.

3) Limiting the similarity score to improve the model: Besides evaluating the generated samples, adding some conditions to the model can also limit the error rate and improve the operation efficiency of the model. The similarity score is used for judging whether the generated word is similar to the original one in the aspect of semantics and part of speech.

Originally, the model is aimed to collect the 50 most similar synonyms no matter what the scores they are. However, some words only have synonyms with low score, which means these synonyms' semantics and part of speech is not close to the original word. Therefore, using Algorithm 2 will delete the words which are of low scores. From the algorithm, it will get the similarity score of every synonym and delete the words that has score lower than 0.7 (default limit parameter).



```
Algorithm 2 Find the synonym whose similarity score is larger than limit
Input: original_sent_info, generated_sent_info, limit == 0.7
Output: can_replace_pos
for i ← 0 to list_length - 1 do
    $S_1$ ← Sentence ( original_sent_info [0...i] )
    $S_2$ ← Sentence ( generated_sent_info [0...i] )
    for j ← 0 to sent_length - 1 do
        $pos_1$ ← $S_1$ [0...j] [0]
        $pos_2$ ← $S_2$ [0...j] [0]
        $score_1$ ← $S_1$ [0...j] [1]
        $score_2$ ← $S_2$ [0...j] [1]
        if $pos_1$ ) == $pos_2$ and $score_2$ ) ¿= limit then
            return true
return false
```

4) Changing stop words to improve the model: Stop word is defined as the words that is less important and has no related meaning to the context [16]. Therefore, when these words are deleted from the sentence, they will not affect the core meaning of the sentence, which is exactly what we expect because deleting these meaningless words will reduce load and improve efficiency. Moreover, in our model, we generate offensive sentences by changing the core words in the sentence. Thus, in the calculation process, these meaningless words should be ignored or deleted.

B. The dataset and models

1) The dataset CoLA: CoLA is one kind of text classification datasets from General Language Understanding Evaluation (GLUE) benchmark [14], which uses a large set of tools to evaluate the performance of natural language understanding tasks. GLUE contains many tasks and all of the tasks are about single sentence or two sentences classification [14]. CoLA is one of the single sentence classifications from GLUE and it is the corpus about the judgments of English acceptability from books and journals [14].

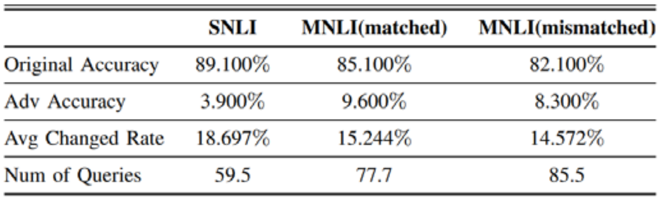
2) The dataset SST-2: Stanford Sentiment Treebank (SST) dataset is a powerful sentiment detection dataset that has abundant resources of training and evaluation [17].

3) The pre-trained model ALBERT: A Lite BERT (ALBERT) is self-supervised pre-trained model similar to BERT. However, ALBERT has a lot of advantages over the original BERT. It is created to solve the issues of memory consumption and training speed and helps deal with the multi-sentence inputs problem [18].

4) The pre-trained model RoBERTa: A Robustly Optimized BERT Pretraining Approach (RoBERTa) is a replication training model of BERT and after evaluating the hyperparameter tuning and set size effects, it becomes more competitive with other models [19].

C. Results

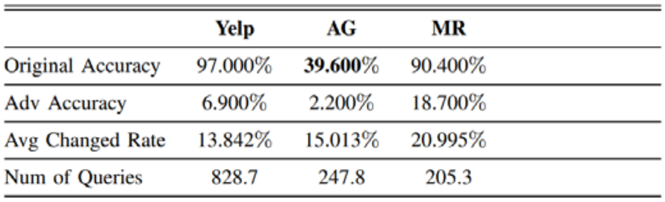
1) Results of implementing the original model: Firstly, we used the original model to attack two tasks with different datasets SNLI, MNLI(matched) and MNLI(mismatched). Specifically, in the task of textual entailment from Table II, after attacked, their accuracy decreased dramatically, from about 85% to 7.2%. When compared with original results, the accuracies, changing rate and number of queries are similar.



	SNLI	MNLI(matched)	MNLI(mismatched)
Original Accuracy	89.100%	85.100%	82.100%
Adv Accuracy	3.900%	9.600%	8.300%
Avg Changed Rate	18.697%	15.244%	14.572%
Num of Queries	59.5	77.7	85.5

TABLE II: Result of attacking SNLI, MNLI (matched and mismatched) in the task of textual entailment (BERT)

Table II is the text classification task. It also showed a powerful attacking performance because it attacked the original accuracy heavily and made it only to around 9.3%. But we got an abnormal original accuracy in the AG dataset. It may be caused by the error of pre-trained model.

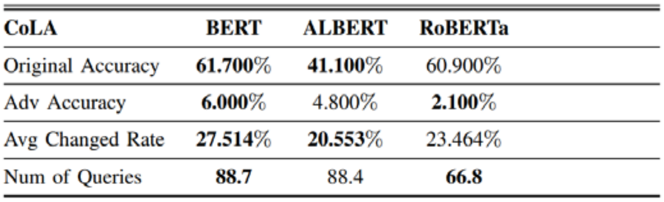


	Yelp	AG	MR
Original Accuracy	97.000%	**39.600%**	90.400%
Adv Accuracy	6.900%	2.200%	18.700%
Avg Changed Rate	13.842%	15.013%	20.995%
Num of Queries	828.7	247.8	205.3

TABLE III: Result of attacking Yelp, (matched and mismatched) in the task of text classification (BERT)

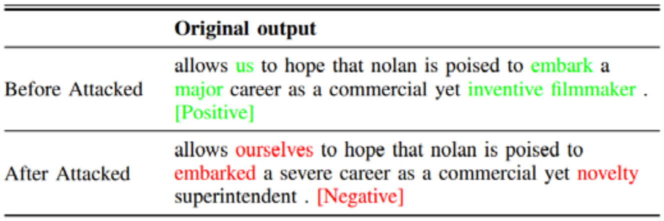
Furthermore, we also made changes to the original experiments. Except form above datasets, we chose a new dataset called CoLA. Additionally, different versions of BERT are also tested. Thus, we got the new results in Table IV. It is clear that after our attack, the accuracies of CoLA-pretrained BERT and ALBERT models decline sharply. Moreover, the average changed rates (how much words are perturbed) are much higher than other datasets. However, the original accuracy is lower than standard level.

2) Results of evaluation algorithm: Like the first example (Grammatical errors), this is one of the common grammatical errors happened in the process of attacking, that is, the tense of verbs. It seems that although there are some judgements to the part of speech of a word, the algorithm cannot choose the tense of verbs correctly in some cases. In the Table V, the word "embarked" is replaced by the word "embarked", which does not follow grammar rules. The ideal replaced word should not only be a synonym of "embark" like "commence", "enter" or "launch", but it should also successfully attack the model.



CoLA	BERT	ALBERT	RoBERTa
Original Accuracy	**61.700%**	**41.100%**	60.900%
Adv Accuracy	**6.000%**	4.800%	**2.100%**
Avg Changed Rate	**27.514%**	**20.553%**	23.464%
Num of Queries	**88.7**	88.4	**66.8**

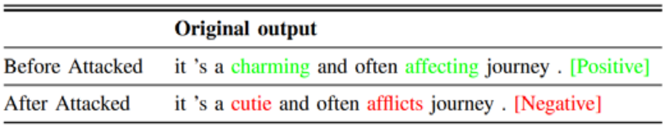
TABLE IV: Result of attacking CoLA (matched and mismatched) with different models in text classification



	Original output
Before Attacked	allows us to hope that nolan is poised to embark a major career as a commercial yet inventive filmmaker . [Positive]
After Attacked	allows ourselves to hope that nolan is poised to embarked a severe career as a commercial yet novelty superintendent . [Negative]

TABLE V: Original generated sample that violates syntax

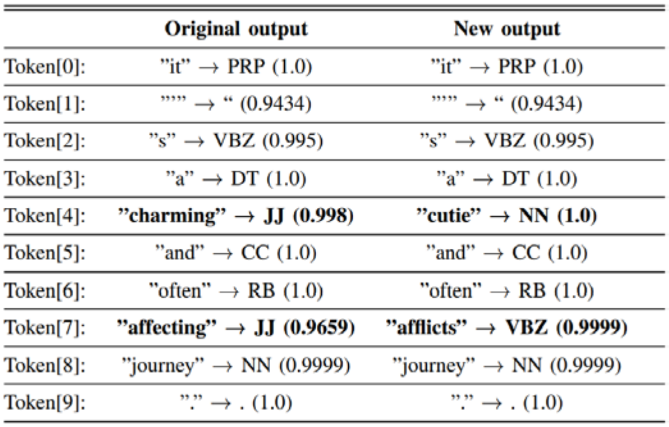
Moreover, there is another generated sample with wrong syntax error.



	Original output
Before Attacked	it 's a charming and often affecting journey . [Positive]
After Attacked	it 's a cutie and often afflicts journey . [Negative]

TABLE VI: Original generated sample that violates syntax

From Table VI, it is clear that this model mistakenly replaces "affecting" with "afflicts", which is also an error about syntax. More specifically, putting the original and generated sentences into the language tool "flair", we can get details in Table VII.

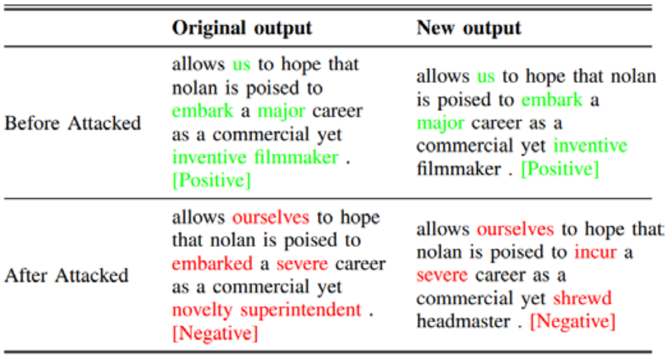


	Original output	New output
Token[0]:	"it" → PRP (1.0)	"it" → PRP (1.0)
Token[1]:	”””” → “ (0.9434)	”””” → “ (0.9434)
Token[2]:	”s” → VBZ (0.995)	”s” → VBZ (0.995)
Token[3]:	”a” → DT (1.0)	”a” → DT (1.0)
Token[4]:	**”charming” → JJ (0.998)**	**”cutie” → NN (1.0)**
Token[5]:	”and” → CC (1.0)	”and” → CC (1.0)
Token[6]:	”often” → RB (1.0)	”often” → RB (1.0)
Token[7]:	**”affecting” → JJ (0.9659)**	**”afflicts” → VBZ (0.9999)**
Token[8]:	”journey” → NN (0.9999)	”journey” → NN (0.9999)
Token[9]:	”.” → . (1.0)	”.” → . (1.0)

TABLE VII: Analysis about the sample with ”flair”

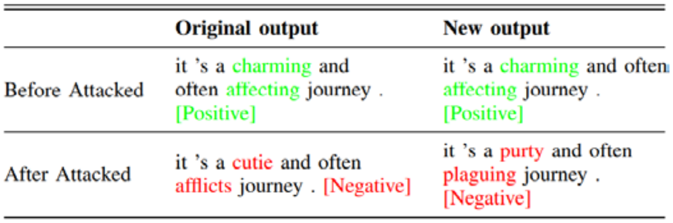
For the fourth token, the word with part of speech "JJ"(Adjective) is changed to that of "NN"(Noun, singular or mass). When it comes to the seventh token, the original part of speech is "JJ"(Adjective). After implementing the model, it is replaced by the word with part of speech "VBZ"(Verb, 3rd person singular present).

The above two samples do not conform to the aims of model design. When looking into the original model, we can find that the errors of part of speech are caused by limited elements in language tag set. In the original language tag set, it used only 12 different language tags. Thus, it is not enough and not possible to deal with other words. Therefore, we need to replace the part of speech checking tool with a more powerful one called "flair" and setting the parameter to “pos-fast”, some part of speech errors can be solved successfully. This new language tool model has 41 different language tags and it can highly improve the rate of errors that happened because of wrong part of speech.



	Original output	New output
Before Attacked	allows us to hope that nolan is poised to embark a major career as a commercial yet inventive filmmaker . [Positive]	allows us to hope that nolan is poised to embark a major career as a commercial yet inventive filmmaker . [Positive]
After Attacked	allows ourselves to hope that nolan is poised to embarked a severe career as a commercial yet novelty superintendent . [Negative]	allows ourselves to hope that nolan is poised to incur a severe career as a commercial yet shrewd headmaster . [Negative]

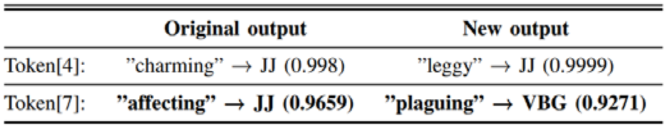
TABLE VIII: Improved generated samples 1



	Original output	New output
Before Attacked	it 's a charming and often affecting journey . [Positive]	it 's a charming and often affecting journey . [Positive]
After Attacked	it 's a cutie and often afflicts journey . [Negative]	it 's a purty and often plaguing journey . [Negative]

TABLE IX: Improved generated samples 2

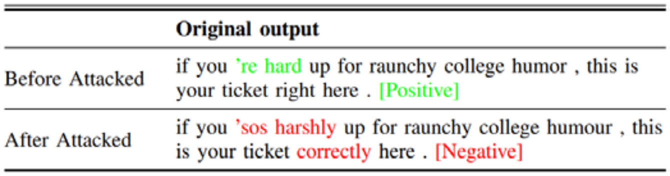
As in Table VIII and Table IX, the wrong tense of verb “embark” is corrected to "incur". As in the next example, the incorrect generated words "charming" is replaced by "purty", and “afflicts” is changed into “plaguing”, which are the synonyms as the original words "charming" and “affecting”. For the second example, when putting the new generated sentence into evaluation algorithm, we can get the correct compared part of speech as in Table X. The algorithm gives that "plaguing" is "VBG"(Verb, gerund or present participle), which can also be used as "JJ"(Adjective).



	Original output	New output
Token[4]:	”charming” → JJ (0.998)	”leggy” → JJ (0.9999)
Token[7]:	**”affecting” → JJ (0.9659)**	**”plaguing” → VBG (0.9271)**

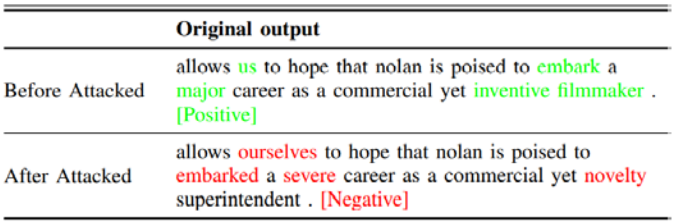
TABLE X: Analysis about improved generated words 1

3) Results of changing stop words: We can also get some bad samples with no relation to part of speech errors as in Table XI and Table XII.



	Original output
Before Attacked	if you 're hard up for raunchy college humor , this is your ticket right here . [Positive]
After Attacked	if you 'sos harshly up for raunchy college humour , this is your ticket correctly here . [Negative]

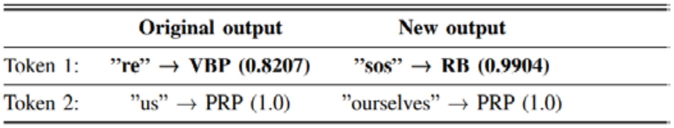
TABLE XI: Generated sample with stop word error 1



	Original output
Before Attacked	allows us to hope that nolan is poised to embark a major career as a commercial yet inventive filmmaker . [Positive]
After Attacked	allows ourselves to hope that nolan is poised to embarked a severe career as a commercial yet novelty superintendent . [Negative]

TABLE XII: Generated sample with stop word error 2

After putting the original and the new generated sentences into the evaluation algorithm. In Table XIII, it illustrates that the part of speech of both original and replaced words. For the token from the first example sentence, they have different part of speech. Moreover, "sos" also does not seem to make any sense. Thus, this replaced word is not correct and it should not be substituted. For another token from second example sentence, although they share the same part of speech, these kinds of words are meaningless for the central meaning of the sentence and also should not be considered in the word importance ranking. Therefore, these two tokens should not be replaced with new adversarial words.



	Original output	New output
Token 1:	**”re” → VBP (0.8207)**	**”sos” → RB (0.9904)**
Token 2:	”us” → PRP (1.0)	”ourselves” → PRP (1.0)

TABLE XIII: Analysis about improved generated words 2

These samples indicate that the model incorrectly turned “ ‘re” into “ ‘sos” and "us" into "ourselves". Actually, in most of cases, we agreed to call these as stop words. It is because that even deleting them will have no effect on the central meaning of the sentence. Firstly, print out the original stop words list. From the list, there is no stop word Table XIV and we need to add these words to the stop words list.

	Example
’ve	you’ve, I’ve ...
’re	you’re ...
’ll	you’ll, I’ll ...
’d	I’d ...
n’t	don’t ...
Special value	us, ourselves ...

TABLE XIV: New added stop words list

This kind of error can be solved by adding such stop words and here is the result after changing the stop word list as in Table XV and Table XVI.

	Original output
Before Attacked	if you ’re hard up for raunchy college humor , this is your ticket right here . [Positive]
After Attacked	if you ’re harshly up for raunchy college humour , this is your ticket correctly here . [Negative]

TABLE XV: Generated sample with changed stop words 1

	Original output
Before Attacked	allows us to hope that nolan is poised to embark a major career as a commercial yet inventive filmmaker . [Positive]
After Attacked	allows us to hope that nolan is poised to embarked a severe career as a commercial yet novelty superintendent . [Negative]

TABLE XVI: Generated sample with changed stop words 2

4) Results of score limiter: In this part, after implementing the new algorithm which limits the similarity score above 0.7(default value), we can get the results in Table XVII. The original output showed that it takes the following words whose similarity scores are below 0.7 into consideration. After improved, the new output indicates that these words are no longer considered, which improves the efficiency.

4) Analyzing of the whole results: After implementing the original model, the Table XVIII shows around 60 percent of generated sentences have the semantical and grammar errors. On the whole, there are 108 errors about wrong used words. It shows that the original model produces low quality of samples. When it comes to the model with new algorithm and stop words list, error sentences reduce to 17 and the total wrong words is much fewer, which is only 23.

Moreover, the most kind of error happened when the model is trying to changed one word into "NN"(Noun, singular or mass). For the new output results, there are only 11 words with such error type. From other hand, "NNP"(Proper noun, singular) is the least error that happened in the original model, while in the new output, "IN"(Preposition or subordinating conjunction), "VBP"(Verb, non-3rd person singular present) and "VBD"(Verb, past tense) take place one time. To conclude, the rate of average sentences with errors in the previous model is around 67%, while the new output with new algorithms produces better results and the error rate reduced to 20%.

Original output	New output
...	...
theatrical [’NN’, 0.5351541042327881]	Non-existent
deft [’JJ’, 0.574834942817688]	Non-existent
zany [’JJ’, 0.5474701523780823]	Non-existent
tempting [’JJ’, 0.4123070240020752]	Non-existent
bewitching [’JJ’, 0.5704071521759033]	Non-existent
beguiling [’JJ’, 0.6152840852737427]	Non-existent
captivating [’JJ’, 0.5463736057281494]	Non-existent
...	...

TABLE XVII: Words with low similarity score

Number of	Original output	New output
sentences with errors	60	**17**
words with errors	108	**23**
most wrong replaced words POS	64 (NN)	**11 (NN)**
least wrong replaced words POS	1 (NNP)	**1 (IN, VBP, VBD)**
Average sentences with errors	67.4%	**20.0%**

TABLE XVIII: Difference between original and new model

III. CONCLUSIONS AND FUTURE WORK

A. Conclusion

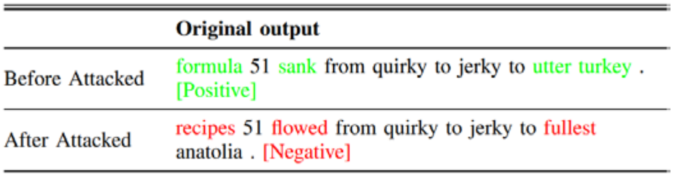
To conclude, all the tasks proposed in the paper have completed, with both implementation and experimental results being obtained. Moreover, improvements about the low-quality generated samples are presented in tables and they decrease the error rate and increase the efficiency of the mode.

Our method is effective and it shows that adversarial attacking models play a great role in increasing the robustness of BERT. Our implementation process and experiments truly reflect the powerful role of EmotionFooler in textual attacking. It can always attack models with an original accuracy of in high level around 90% to a much lower level less than 10%. Also, the improvement part fixes the disadvantages of the original model and increase the quality of generated samples.

B. Future Work

One direction is to experiment and to create a similar natural language processing model and toolbox in a trusted environment to obtain a better result.

Although the errors of verb tense do not follow grammar rules, semantic of specific candidate words in a sentence seems unnatural and low similarity words are selected are solved, we can still further improve the model in candidate words. For instance, the terminology problem in Table XIX.



	Original output
Before Attacked	formula 51 sank from quirky to jerky to utter turkey . [Positive]
After Attacked	recipes 51 flowed from quirky to jerky to fullest anatolia . [Negative]

TABLE XIX: Error of terminology

Sometimes the model will replace this kind of terminology or special word. In this case, new model replaces the "formula 51" into "recipes 51", which confuses human because in our language, "formula 51" is the name of one kind of car. Thus, this kind of generated sentences are not natural and it should not happen in our model. Besides, we may also need to pay attention to replace words with different tense or even phrases, which premise is that adversarial samples should always be reasonable in sentence meanings and obey grammar rules.

IV. ACKNOWLEDGMENT

This work is supported by the Xian Jiaotong-Liverpool University (XJTLU) AI University Research Centre, Jiangsu (Provincial) Data Science and Cognitive Computational Engineering Research Centre at XJTLU; and research funding: XJTLU-REF-21-01-002.

REFERENCES

[1] T. Roth, Y. Gao, A. Abuadbba, S. Nepal, and W. Liu, “Token-modification adversarial attacks for natural language processing: A survey,” arXiv preprint arXiv:2103.00676, 2021.

[2] D. Jin, Z. Jin, J. T. Zhou, and P. Szolovits, “Is bert really robust? a strong baseline for natural language attack on text classification and entailment,” in Proceedings of the AAAI conference on artificial intelligence, vol. 34, no. 05, 2020, pp. 8018–8025.

[3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” arXiv preprint arXiv:1810.04805, 2018.

[4] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[5] Y. Zhang and B. Wallace, “A sensitivity analysis of (and practitioners’ guide to) convolutional neural networks for sentence classification,” arXiv preprint arXiv:1510.03820, 2015.

[6] A. Chakraborty, M. Alam, V. Dey, A. Chattopadhyay, and D. Mukhopadhyay, “Adversarial attacks and defences: a survey (2018),” arXiv preprint arXiv:1810.00069, 1810.

[7] W. E. Zhang, Q. Z. Sheng, A. Alhazmi, and C. Li, “Adversarial attacks on deep learning models in natural language processing: A survey,” ACM Transactions on Intelligent Systems and Technology (TIST), vol. 11, no. 3, pp. 1–41, 2020.

[8] I. J. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and harnessing adversarial examples,” arXiv preprint arXiv:1412.6572, 2014.

[9] J. Ebrahimi, A. Rao, D. Lowd, and D. Dou, “Hotflip: White-box adversarial examples for text classification,” arXiv preprint arXiv:1712.06751, 2017.

[10] J. Li, S. Ji, T. Du, B. Li, and T. Wang, “Textbugger: Generating adversarial text against real-world applications,” arXiv preprint arXiv:1812.05271, 2018.

[11] M. Behjati, S.-M. Moosavi-Dezfooli, M. S. Baghshah, and P. Frossard, “Universal adversarial attacks on text classifiers,” in ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019, pp. 7345–7349.

[12] Y. Zang, F. Qi, C. Yang, Z. Liu, M. Zhang, Q. Liu, and M. Sun, “Word-level textual adversarial attacking as combinatorial optimization,” arXiv preprint arXiv:1910.12196, 2019.

[13] S. Garg and G. Ramakrishnan, “Bae: Bert-based adversarial examples for text classification,” arXiv preprint arXiv:2004.01970, 2020.

[14] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman, “Glue: A multi-task benchmark and analysis platform for natural language understanding,” arXiv preprint arXiv:1804.07461, 2018.

[15] “Huggingface,“ https://huggingface.co/, 2022.

[16] A. Alajmi, E. M. Saad, and R. Darwish, “Toward an arabic stop-words list generation,” International Journal of Computer Applications, vol. 46, no. 8, pp. 8–13, 2012.

[17] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, “Recursive deep models for semantic compositionality over a sentiment treebank,” in Proceedings of the 2013 conference on empirical methods in natural language processing, 2013, pp. 1631–1642.

[18] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, “Albert: A lite bert for self-supervised learning of language representations,” arXiv preprint arXiv:1909.11942, 2019.

[19] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, “Roberta: A robustly optimized bert pretraining approach,” arXiv preprint arXiv:1907.11692, 2019.

Moving towards sustainable mobility: Examining the determinants of electric vehicles purchase intention in India

Jitender Kumar Atri, Woon Kian Chong and Muniza Askari

***Abstract*: India's battery electric vehicle (BEV) market growth has been exponential. This research investigates the factors influencing the purchase intention (PI) of EV customers drawing upon the extended theory of planned behavior (TPB). We employ a two-phase study starting with a pilot study and the main study. The research confirmed that range and infrastructure readiness, price value, emotional value and environmental concern all significantly affected attitude. The study also established that environmental concern, attitude, subjective norms and perceived behavioral control significantly impacted purchase intention. The study additionally concluded that safety does not significantly affect attitude. Our results suggest a mix of push (government policies, technology improvement, infrastructure, etc.) and pull (customer's purchase intention) to accelerate the BEV market growth.**

***Keywords:* Electric Vehicles, India, Purchase Intention, Theory of Planned Behavior**

I. INTRODUCTION

The transport sector is amongst the largest energy-consuming sectors. It is globally overly dependent on hydrocarbon-based fossil fuels. The industry is also a significant source of Green House Gas (GHG) emissions and accounts for 24 % of total global energy-related carbon dioxide (CO_2) emissions [1]. The transport sector of India is the third most GHG emitting industry, of which the road transport sector is the major contributor. Out of the total CO_2 emissions in India, it was reported that 13% come from the transport sector [2]. Further, due to increased energy needs, crude oil import for India has risen ten times since 1990 [3]. Such a high dependency on energy sources on imports severely affects national energy security [4].

The global automotive industry is exploring alternatives to internal combustion engines (ICE). Electrification is one of the solutions to address the increasing levels of vehicle pollution. Electrification of vehicles has several benefits, like higher efficiency and lower air pollution, thereby reducing CO_2 [5]. Technological advancements supported by government policies and regulations has helped the BEV market grow. [6] reported that EV sales doubled in 2021 compared to 2020, with global sales of 6.6 million.

To create momentum for the adoption of EVs in India, the government is working toward providing tax incentives and applying stringent targets for carbon emissions via Corporate Average Fuel Consumption (CAFC). These initiatives will also help to enhance infrastructure and support Research & Development (R&D) for technological advancement [7]. These steps and policies will push toward creating an environment that is more acceptable to EVs. In addition, improved technology will create a pull by the customers with enhanced infrastructure, range, and price reduction. Marketing also needs to create positive attitude of customers toward EVs and inform the potential customers of environment benefits of BEVs.

All authors are with the SP Jain School of Global Management
(email: Jitenderkumar.db1804007@spjain.org, tristan.chong@spjain.org, Muniza.askari@spjain.org).

II. LITERATURE REVIEW

THEORY OF PLANNED BEHAVIOR (TPB)

The Theory of Planned Behaviour (TPB) has been an adequate and influential model in explaining or predicting behaviour intentions [8-10]. Moreover, it has successfully attracted wide application and empirical support for several pro-environmental behaviours, as illustrated below:

- Attitude refers to individuals' positive or negative evaluation of performing a behaviour. Attitude results from behavioural beliefs and outcome evaluations. Behavioural belief refers to the unique idea about the consequences of engaging in a particular behaviour. While outcome evaluation refers to the corresponding favourable or unfavourable judgment about the possible consequences of the behaviour [10].
- Subjective norms (SN) represent the social pressure from the reference group members to act on a given behaviour. Subjective norm is an outcome of normative belief and motivation to comply. Normative belief refers to an individual perception of how others (those who are significant to the individual) would like one to behave in a particular situation. Whereas motivation to comply refers to the individual desire to adhere to the opinion of significant others [10].
- Perceived behavioural control (PBC) concerns the perceived ease or difficulty of performing a behaviour. PBC is an outcome of control beliefs and perceived power. Control belief can be defined as the individual's belief towards the presence of certain factors that may facilitate or impede the performance of a particular behaviour (e.g., time, money & opportunity). On the other hand, perceived power refers to the personal evaluation of the impact of these factors in facilitating or impeding a particular behaviour [10].

- Behavioural intention indicates an individual's readiness to perform a given behaviour. It is assumed to be an immediate antecedent of behaviour [10]. The more favourable the attitude towards behaviour, the better the subjective norm, and the greater the perceived behavioural control, the stronger the individual's intention to perform the behaviour will be. The intention indicates " how hard people are willing to try, of how much effort they are planning to exert, to perform the behaviour." [11]

EXTENDED THEORY OF PLANNED BEHAVIOR

"The theory of planned behavior is, in principle, open to the inclusion of additional predictors if it can be shown that they capture a significant proportion of the variance in intention or behavior after the theory's current variables have been taken into account." [10]. Although it is well known that TPB assumes that intention to perform the behavior is derived from attitude, subjective norm, and PBC, researchers in the past advocated for domain-specific factors, which are not included in this model. Perceived value and willingness to pay a premium (WPP) were added along with Attitude, Subjective Norm and PBC for measuring consumers' green purchase intention [9]. Researchers have also considered price and emotional value, as it plays a vital role in green purchase decisions as consumers will not compromise on the functional benefit of the product just for the sake of the environment. Therefore, understanding consumers' value of green products is crucial. Further, willingness to pay a premium was considered as the high price of eco-friendly products is still an issue for price-sensitive Indian consumers [9]. Based on the above, we anticipated that the following factors are essential, as illustrated in fig. 1 below:

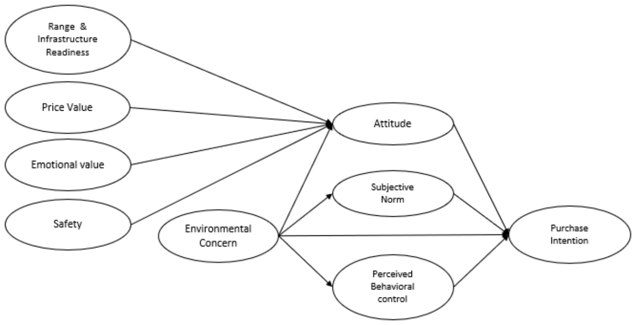


Fig 1. Conceptual Framework

RANGE AND INFRASTRUCTURE READINESS (RI)

The driving range of BEVs has long been considered a significant barrier to the acceptance of electric mobility. Due to the limited range of BEVs, drivers feel stressed about becoming stranded if the battery charge is depleted, known as range stress. However, higher levels of trust in range estimates lead to lower range stress and higher acceptance of BEVs [12]. Effects of range anxiety can be significant but are reduced with access to additional charging infrastructure [13].

High acquisition costs and short driving ranges are the main factors that impede EV diffusion [14]. In addition, EVs suffer from a short travel distance on a battery charge, a lack of charging infrastructure, and long charging times, collectively called charging risk, which are the primary reasons consumers are reluctant to adopt them [15]. On the other hand, fast-charging infrastructure will benefit and facilitate long-range drives for EVs, which may be crucial to pushing the market penetration of EVs. Thus, Infrastructure readiness plays an essential role in increasing market penetration and public acceptance [7].

PRICE VALUE (PV)

Price value can be defined as consumers' mental tradeoff between the perceived benefits of the action and the cost of using them [16]. The price value is positive when the advantages of using technology are perceived to be greater than the cost.

PV predicts behavioral intention to use technology [17]. Individuals consider vehicle costs and performance characteristics as important factors when choosing their next vehicle. Also, they were attracted to "tax-free purchase" incentives and vehicles with significantly reduced emission levels [18]. BEVs lack economies of scale and so they are relatively expensive. The cost of replacing BEV batteries is also a burden that ICE vehicles do not impose. Therefore, price and battery cost negatively affect the perceived value of BEVs [15]. Willingness to pay a premium (WPP) was not found to influence green purchase intention significantly, implying that the consumers in India are more sensitive toward price value [9].

EMOTIONAL VALUE (EV)

Emotions have been added to decision models such as the TPB for various products and issues and have been proven to improve the model's predictability. Emotions enhance the explanatory power of the TPB in predicting intentions for cornea donations [19]. Sentiments towards the BEVs and thoughtful emotions toward car driving strongly affect usage intention [20]. Emphasizing the importance of emotional value, [21] commented that if companies neglect to appeal to consumers' emotions, even an excellent technology product could go wasted.

ENVIRONMENTAL CONCERN (EC)

EC has a moderating effect on PI through Attitude, SN and PBC, in addition to having a direct effect on PI [22]. Concern for the environment is significantly related to consumer behavior, including purchasing intentions [23].

SAFETY

[24] in their study stated that Battery stability is a major safety concern related to electric vehicles that can considerably affect customer attitude. Though with improved technology and advancement of Battery Management System (BMS), the current Lithium-ion batteries have proved to have enhanced safety measures related to fires. But the technology is still advancing, and we still see thermal incidences in BEVs.

PURCHASE INTENTION (PI)

Purchase Intention of environmentally friendly products can be defined as "the likelihood that a consumer would buy a particular product resulting from his or her environmental needs" [25].

III. PILOT STUDY

In the pilot and main study, the researcher considered a survey strategy to collect the data on BEVs PI, with a questionnaire made available over the internet. In the pilot phase, 115 inputs were collected to check the indicator variable loading on their latent variables. Based on the measurement model of the pilot study, two additional indicator variables were added before the main study.

IV. MAIN STUDY

During the main study, 1008 inputs were received, of which 67 respondents already had a BEV and were omitted; additionally, 157 respondents did not complete the whole questionnaire and were thus excluded from the analysis. Finally, 784 inputs were coded for final analysis and model testing. ADANCO 2.3.1 was used to run the model and analyze the data. Fig.2 shows the output SEM model (Electric Vehicle Purchase Intention (EVPI)) of this study. Table 5 gives the details of the path coefficients, and table 6 for the R^2 values shown in the model.

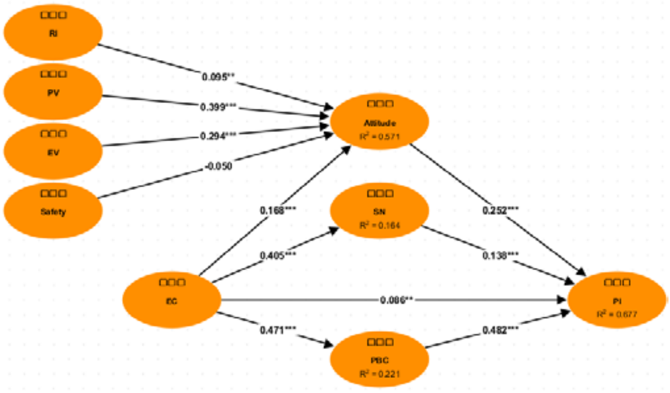


Fig. 2. The main study SEM model - EVPI

Main study sample size and composition

Variable	***Category***	***Frequency***	***Percentage***
Gender	Male	624	80
	Female	160	20
Age	18-25 years	92	12
	26-30 years	80	10
	31-40 years	270	34
	41-50 years	253	32
	Above 51	89	11
Education	Grade 10	1	0
	Grade 12	17	2
	Graduation	410	52
	Post Grad	344	44
	Phd	12	2
Income	Less than 5 lacs	72	9
	5 - 10 lacs	140	18
	10 - 20 lacs	186	24
	20 - 30 lacs	175	22
	>30 lacs	211	27

Table 1. Main study sample composition

Construct Reliability

Construct reliability relates to the measured consistency of an observed indicator towards the construct that it is measuring. Dillon-Goldstein's rho is a better reliability measure than Cronbach's alpha in Structural Equation Modeling since it is based on the loadings rather than the correlations observed between the observed variables [26]. Reliability values above 0.7 confirms construct reliability [27]. Table 2 values suggest construct reliability of the model, with all values above 0.7.

Construct	Jöreskog's rho (ρ_c)
RI	0.8132
PV	0.8360
EV	0.8443
EC	0.7946
Safety	0.8488
Attitude	0.8934
SN	0.8707
PBC	0.8441
PI	0.9339

Table 2. Construct reliability for main study

Convergent Validity

Convergent validity signifies that a set of indicators represents the same underlying construct [28]. As a parameter, convergent validity ascertains the degree to which two measures of constructs that should theoretically be related are, in fact, related. Average variance extracted (AVE) figures have been analyzed to test the convergent validity of the model. The satisfactory threshold for this measurement is 0.5 [27]. Table 3 shows the AVE values, all above 0.5, confirming the convergent validity of the model.

Construct	Average variance extracted (AVE)
RI	0.5236
PV	0.5648
EV	0.7314
EC	0.5713
Safety	0.7428
Attitude	0.7372
SN	0.6976
PBC	0.7307
PI	0.8249

Table 3. Convergent Validity for main study

Discriminant Validity

Discriminant validity means that two conceptually different constructs must also differ statistically or that a latent variable shares more variance with its assigned indicators than another latent variable in the structural model [29].

In statistical terms, the AVE of each latent construct should be greater than the highest squared correlation with any other latent construct. Therefore, ADANCO output includes a table called "Discriminant Validity: Fornell-Larcker Criterion," containing the reflective constructs' average variance extracted in its main diagonal and the squared inter-construct correlations in the lower triangle. Discriminant validity is regarded as given if the highest absolute value of each row and each column is found in the main diagonal. Table 4 shows that using the Fornell-Larker criterion, discriminant validity is confirmed.

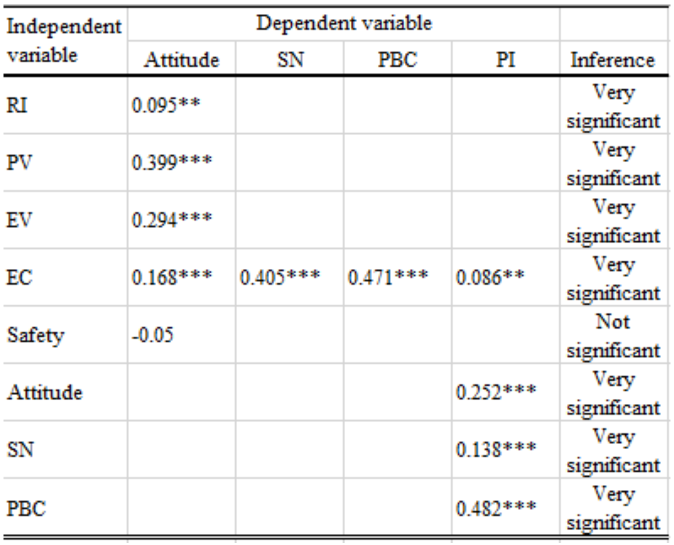
	RI	PV	EV	EC	Saf.	Att.	SN	PBC	PI
RI	0.52								
PV	0.27	0.56							
EV	0.13	0.25	0.73						
EC	0.09	0.16	0.18	0.57					
Saf.	0.00	0.00	0.01	0.01	0.74				
Att.	0.21	0.44	0.36	0.23	0.01	0.74			
SN	0.12	0.16	0.21	0.16	0.00	0.28	0.70		
PBC	0.16	0.28	0.32	0.22	0.01	0.52	0.29	0.73	
PI	0.16	0.28	0.28	0.24	0.00	0.51	0.32	0.61	0.82

Squared correlations; AVE in the diagonal

Table 4. Discriminant Validity for main study

Path Coefficient

The path coefficients are standardized regression coefficients (beta values). A path coefficient quantifies the direct effect of an independent variable on a dependent variable. For example, path coefficients are interpreted as the increase in the dependent variable if the independent variable were increased by one standard deviation, and all the other independent variables in the equation remained constant [28]. Table 5 shows the path coefficients between the latent variables and the p-value indicator. Price value (PV) is the most critical factor that affects attitude, followed by emotional value (EV) and environmental concern (EC). Even though range and infrastructure significantly affect attitude, the effect was the least. PBC has the highest effect on Purchase Intention (PI), followed by attitude and subjective norms. All the effects were significant other than safety-> attitude.



Independent variable	Dependent variable				Inference
	Attitude	SN	PBC	PI	
RI	0.095**				Very significant
PV	0.399***				Very significant
EV	0.294***				Very significant
EC	0.168***	0.405***	0.471***	0.086**	Very significant
Safety	-0.05				Not significant
Attitude				0.252***	Very significant
SN				0.138***	Very significant
PBC				0.482***	Very significant

Table 5. Path Coefficients

Here, ** Signifies p<0.01

*** Signifies p<0.001

Coefficient of Determination

R^2, or the coefficient of determination, measures the variance explained in each of the endogenous constructs and is, therefore, a measure of the model's explanatory power [30]. Table 6 suggests that 0.68 or 68% of the variance of PI is explained by the contributing factors included as antecedents in this model.

Construct	R^2	Adjusted R^2
Attitude	0.57	0.57
SN	0.16	0.16
PBC	0.22	0.22
PI	0.68	0.68

Table 6. R^2 values for latent variables

Effect	**Original coefficient**	**t-value**	**p-value (2-sided)**	**p-value (1-sided)**	**Supported**
Range & Inf-> Attitude	0.09	3.26	0.00	0.00	Yes
Range & Inf -> PI	0.02	2.93	0.00	0.00	Yes
PV -> Attitude	0.40	12.87	0.00	0.00	Yes
PV -> PI	0.10	5.60	0.00	0.00	Yes
EV -> Attitude	0.29	9.82	0.00	0.00	Yes
EV -> PI	0.07	5.58	0.00	0.00	Yes
EC -> Attitude	0.17	5.68	0.00	0.00	Yes
EC -> SN	0.41	11.89	0.00	0.00	Yes
EC -> PBC	0.47	14.09	0.00	0.00	Yes
EC -> PI	0.41	12.63	0.00	0.00	Yes
Safety -> Attitude	-0.05	-1.58	0.11	0.06	No
Safety -> PI	-0.01	-1.51	0.13	0.07	No
Attitude -> PI	0.25	6.60	0.00	0.00	Yes
SN -> PI	0.14	4.90	0.00	0.00	Yes
PBC -> PI	0.48	13.07	0.00	0.00	Yes

Table 7. Total effect inference

V. DISCUSSION AND IMPLICATIONS

This study explores the factors that influence the PI of BEV in India. Developing on extended TPB, in addition to Attitude, Subjective Norm, and PBC, the framework was expanded to other factors such as range confidence, price value, emotional value, environmental concern and safety. The EVPI model was validated, and all other effects were significant other than safety. The price value was the most crucial factor affecting attitude, followed by the emotional value. Range and infrastructure readiness had the most negligible effect on attitude. PBC had the maximum effect on PI, followed by attitude and subjective norms.

The price value is the most critical factor, and actions to reduce the initial cost and overall cost of ownership are essential. The government of India has already initiated many reforms, subsidies, and regulations to promote nationwide BEV volumes[7]. Additionally, to promote R&D and develop localized products, Production Linked Incentive(PLI) scheme has been rolled out by the government in 2022. The technological advancement and subsidies will help address the most critical factor of price value (cost of acquisition, cost of ownership). To address the emotional value of customers, suitable marketing strategies need to be implemented to appeal to the emotions of potential customers. These customers should feel proud of owning BEVs; a distinct differentiation of EVs from other vehicles will help towards this aspect. Furthermore, to bring more awareness to the environmental benefits of BEVs, suitable communications need to be put in place by the government and OEMs. Finally, the range needs to meet the minimum expectations of the customers in addition to charging stations at home and workplaces for daily use customers and 150-200 km between major cities for intercity commuters.

VI. CONCLUSION

India has been witnessing exponential growth in BEV during the last two years. However, work must be done at multiple levels for this change to be smooth and faster. With price value as the most critical factor for success, technological innovation is significant in reducing the cost of acquisition and the overall cost of ownership. To further improve the buying proposition of BEVs, government subsidies towards BEVs and stringent regulations towards ICE vehicles need to continue. A consistent, reliable driving range of at least 250 km on a full charge is critical. The BEV customer expects a charging facility at home or the workplace. For intercity commutes, fast charging at distances of 200 Kms from significant cities is essential to charge up to 80% in 30-45 mins. All this can be successful only if abundant green electricity sources are available across the country. To help increase R&D and provide options to potential customers, government initiatives like the PLI will significantly boost OEMs to invest in BEVs. Standardization of chargers across different vehicles will help the utility of any charging facility. An industrywide collaboration will help to standardize charging facilities across manufacturers. To educate customers about BEV benefits, environmental benefits must be communicated widely via all communication channels, including the internet and social media. First-mover anxiety can be reduced as more players come and people start experiencing BEVs. Improved battery technology will not only improve reliability and range confidence but also drive away the fear of safety. Experience will help people overcome range anxiety, and a further improvement in battery technology and improved infrastructure will help strengthen the BEV market.

There is a lack of communication regarding the awareness of BEV, and that's why the attitude towards BEVs hasn't yet improved. Appropriate communication on the benefits of BEVs on the environment needs to be created to motivate potential customers to buy BEVs. In addition, there is a lack of visibility of BEVs on the road. With limited options and anxiety about new technology, the BEV market in India is yet to catch the pace. Government policies to encourage BEVs and regulate ICE vehicles and subsidies must continue to make BEVs lucrative for manufacturers and customers.

REFERENCES

[1] IEA (2020), Tracking Transport 2020, IEA, Paris Retrieved from https://www.iea.org/reports/tracking-transport-2020

[2] Ministry of Environment & Forests Government of India (2010). INCCA: Indian Network for Climate Change Assessment. Retrieved from

[3] http://www.indiaenvironmentportal.org.in/files/fin-rpt-incca.pdf

[4] IEA Oil Information (2022) Retrieved from https://www.iea.org/data-and-statistics/data-product/oil-information)

[5] Dhar, S., Pathak, M., & Shukla, P. R. (2017). Electric vehicles and India's low carbon passenger transport: a long-term co-benefits assessment. Journal of Cleaner Production, 146, 139-148.

[6] Okada, T., Tamaki, T., & Managi, S. (2019). Effect of environmental awareness on purchase intention and satisfaction pertaining to electric vehicles in Japan. Transportation Research Part D: Transport and Environment, 67, 503-513s

[7] Global EV Outlook (2022). Retrieved from https://www.iea.org/reports/global-ev-outlook-2022

[8] Mishra, S., & Malhotra, G. (2019). Is India Ready for e-Mobility? An Exploratory Study to Understand e-Vehicles Purchase Intention. Theoretical Economics Letters, 9(2), 376-391.

[9] Arli, D., Tan, L. P., Tjiptono, F., & Yang, L. (2018). Exploring consumers' purchase intention towards green products in an emerging market: The role of consumers' perceived readiness. International Journal of Consumer Studies, 42(4), 389–401

[10] Yadav, R., & Pathak, G. S. (2017). Determinants of consumers' green purchase behavior in a developing nation: Applying and extending the theory of planned behavior. Ecological economics, 134, 114-122.

[11] Ajzen, I. (1991). The theory of planned behavior. Organizational behavior and human decision processes, 50(2), 179-211.

[12] Ajzen, I. (2002). Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior 1. Journal of applied social psychology, 32(4), 665-683.

[13] Nastjuk, I., Werner, J., Marrone, M., & Kolbe, L. M. (2018). Inaccuracy versus volatility–Which is the lesser evil in battery electric vehicles?. Transportation research part F: traffic psychology and behaviour, 58, 855-870.

[14] Neubauer, J., & Wood, E. (2014). The impact of range anxiety and home, workplace, and public charging infrastructure on simulated battery electric vehicle lifetime utility. Journal of power sources, 257, 12-20.

[15] Degirmenci, K., & Breitner, M. H. (2017). Consumer purchase intentions for electric vehicles: is green more important than price and range?. Transportation Research Part D: Transport and Environment, 51, 250-260.

[16] Kim, M. K., Oh, J., Park, J. H., & Joo, C. (2018). Perceived value and adoption intention for electric vehicles in Korea: Moderating effects of environmental traits and government supports. Energy, 159, 799-809

[17] Dodds, W. B., Monroe, K. B., & Grewal, D. (1991). Effects of price, brand, and store information on buyers' product evaluations. Journal of marketing research, 28(3), 307-319

[18] Venkatesh, V., Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS quarterly, 157-178.

[19] Potoglou, D., & Kanaroglou, P. S. (2007). Household demand and willingness to pay for clean vehicles. Transportation Research Part D: Transport and Environment, 12(4), 264-274.

[20] Bae, H. S. (2008). Entertainment-education and recruitment of cornea donors: The role of emotion and issue involvement. Journal of health communication, 13(1), 20-36.

[21] Moons, I., & De Pelsmacker, P. (2015). An extended decomposed theory of planned behavior to predict the usage intention of the electric vehicle: A multi-group comparison. Sustainability, 7(5), 6212-6245.

[22] Kato, T. (2021). Functional value vs emotional value: a comparative study of the values that contribute to a preference for a corporate brand. International Journal of Information Management Data Insights, 1(2), 100024.

[23] Paul, J., Modi, A., & Patel, J. (2016). Predicting green product consumption using theory of planned behavior and reasoned action. Journal of retailing and consumer services, 29, 123-134.

[24] Lai, I. K., Liu, Y., Sun, X., Zhang, H., & Xu, W. (2015). Factors influencing the behavioural intention towards full electric vehicles: An empirical study in Macau. Sustainability, 7(9), 12564-12585.

[25] Lee, J., & Cho, Y. C. (2021). Fostering Attitudes and Customer Satisfaction for Sustainability by Electric Car-Sharing. The Journal of Industrial Distribution & Business, 12(5), 37-46.

[26] Chen, Y. S., & Chang, C. H. (2012). Enhance green purchase intentions: The roles of green perceived value, green perceived risk, and green trust. Management Decision.

[27] Demo, G., Neiva, E. R., Nunes, I., & Rozzett, K. (2012). Human resources management policies and practices scale (HRMPPS): Exploratory and confirmatory factor analysis. BAR-Brazilian Administration Review, 9, 395-420.

[28] Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. Journal of Marketing Theory and Practice, 19(2), 139–152

[29] Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. In New challenges to international marketing. Emerald Group Publishing Limited.

[30] Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of marketing research, 18(1), 39-50.

[31] Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. European business review, 31(1), 2-24.

Class Token as a Powerful Assistance for Transformer Pretraining

Jingyang Min, Erick Purwanto*, and Su Yang

Abstract— **This paper demonstrates that the class token could be a powerful aid for the Transformer backbone pretraining in computer vision. Specifically, this pretraining task was conducted on a contrastive learning approach, which is the image pair prediction. In the case of conventional contrastive self-supervised learning for binary classifications, the contrastive images would be fed in pairs into the model backbone for training. This work proposes to compute the loss from each class token, which is then summed with the contrastive loss during the pretraining step to obtain an improved prediction. By involving the class token as an auxiliary signal in the pretraining step, the linear evaluation result improved approximately 14%, which is considerably higher than using the conventional training scheme.**

Index Terms— **Learning Representations, Computer Vision, Deep Learning, Contrastive Learning, Pre-training, Transformer.**

I. INTRODUCTION

Nowadays the deep neural networks can admittedly achieve good performance on different tasks, but the time-consuming issue during model training remains one of its critical limitations. Pretraining is considered a promising solution to alleviate this drawback, which in effect, delegating the long-time training and computational expensive requirements to the institutes or companies with enough resources. However, the expense of labelling data remains a challenge. To address this problem, Self-Supervised Learning (SSL), which excludes the need for labels on data, has become a popular theme of research. One particularly promising approach of SSL, namely the Contrastive Self-Supervised Learning (CSL), has attracted much attention in the community in particular.

CSL utilizes the benefits of representation learning from contrastive learning, coupled with the advantages of the labelling exclusion, has been reported with good experiment results for image classifications. Many different network architectures could be adopted as the starting point after CSL pretraining, the Transformer backbone models are the popular candidates. One appealing feature of these Transformer backbone models is they could generate informative embeddings on each corresponding token position. In this paper we present that within the CSL scheme, the information from the class token as a powerful auxiliary signal could also be leveraged to improve the performance of image classification.

II. RELATED WORK

2.1. Contrastive Self-Supervised Learning

Traditional contrastive learning methods are commonly supervised, while some more advanced them could be unsupervised by integrating with SSL [1]. Although SSL has been deployed with a number of applications, such as predicting relative location, solving jigsaw puzzles, and colorizing images, none of them achieved decent performance in practice without contrastive learning [2]. Several CSL methods have achieved competitive performance, such as Augmented Multiscale Deep InfoMax (AMDIM), Contrastive Predictive Coding (CPC), and A Simple Framework for Contrastive Learning of Visual Representations (SimCLR) [3]. These methods first generate multiple image pairs via data augmentation, then feed these pairs to a specific encoder (a different variance of ResNet) to extract the features to obtain the final representation of each image [2]. They use modified forms of Negative Contrastive Estimation (NCE) loss to update parameters in the encoder to achieve good starting point for down-stream tasks, such as classification or object detection tasks [2].

It is generally expected that CSL methods could save manually labelling expense and avoid mislabeled data. As a model could learn the semantic information from designed tasks directly (bypass the incorrectness of labels), a good starting point of models is also more likely to reduce fine-tuning time and examples for adjusting extracted representations in down-stream tasks. From the analysis of experiment results published by M. Caron et al. [3], AMDIM, CPC, and SimCLR still exist noticeable performance gap compared with the supervised learning approaches on most down-stream tasks, and all of them only conduct experiments on CNN backbone encoder instead of Transformer backbone. An SSL approach named Momentum Contrast (MoCo) iterated 3 versions to gradually release the limitation of encoder architecture and improve the performance in down-stream tasks [4, 5, 6]. In addition to trivial encoder such as a single ResNet backbone, the original MoCo utilized an CNN encoder with a momentum encoder, which contains weighted average parameter values accumulated from the CNN encoder branch in a long iteration span, to train the CNN encoder contrastively [4]. Consequently, all these CSL methods are capable of learning robust representations for images.

Although there are several CSL methods that could perform pretrain tasks on deep neural networks, there still exist some

Jingyang Min and Erick Purwanto are with the School of Advanced Technology, Xi'an Jiaotong-Liverpool University (XJTLU), Suzhou, China. Su Yang is with the Department of Computer Science, Faculty of Science and Engineering, Swansea University, United Kingdom (email: jingyang.min18@student.xjtlu.edu.cn, erick.purwanto@xjtlu.edu.cn, su.yang@swansea.ac.uk).

potential drawbacks, such as infinite number of negative image pairs for each original reference image. Therefore, some advanced SSL methods adopted positive image pairs only to train the encoder, including Bootstrap Your Own Latent (BYOL), Swapping Assignments between Views (SwAV), and self-**di**stillation with **no** labels (DINO) [7, 3, 8]. BYOL adopted asymmetric branches named online and target networks to obtain representations for images in positive pairs [7]. Images would be encoded and projected on each branch, then the image presentation came from the online network would be projected further by a predictor [7]. The loss between the prediction and the projection came from the target network would be computed [7]. DINO adopted a similar networks organization structure compared with BYOL, but it has generalized the training method to Transformer architecture and achieved state-of-the-art performance on several down-stream tasks [8]. SwAV implemented SSL with a different concept compared with all SSL methods illustrated before, it will map the computed prototype vectors on spherical surface and assign each vector to the nearest cluster, then swap assignment vectors obtained from corresponding network branches and predict the swapped vector from original branch [3].

All of these methods only adopted data augmentation pipeline to generate positive pairs, then feed generated images into encoder branches with different strategies applied to avoid collapsing, which would happen when there are no negative pairs that exist in SSL process. In conventional CSL methods and the BYOL method, there are often two branches for encoder to learn robust representation of images. However, SwAV introduced the multi-crop trick to generate multiple image patches from the original reference image and trained more robust encoder with multiple parallel branches simultaneously [3]. Therefore, since only positive image pairs were required in these methods, BYOL, DINO, and SwAV could be considered as efficient SSL pretraining schemes.

2.2. Masked Image Modeling Methods

CNN and Transformer could be used for the pretraining of SSL. From previous study of SSL methods, it is reported that most of them performed experiments on CNN architecture. There exist different training schemes rely on Transformer which cannot be categorized as one of the previously illustrated SSL categories, which is the Masked Image Modeling (MIM). Three major representatives in this SSL strategy, including BERT Pre-Training of Image Transformers (BEIT), Masked Autoencoder (MAE), and a Simple Framework for Masked Image Modeling (SimMIM) [9, 10, 11].

In the BEIT, they trained a tokenizer and decoder before pretraining of the Vision Transformer (ViT) encoder, then use the MIM Head to predicate visual tokens at the output of masked image patches [9]. In the Masked Autoencoder (MAE), the Transformer backbone would be trained in an encoder-decoder structure with 75% random masked regions of original images, and the autoencoder would try to reconstruct not masked original images correctly [10]. In the SimMIM, they utilized a similar masking strategy in the MAE except the Transformer backbone was changed to Swin Transformer and adopted a fully connected layer head to predicate masked patches with the criterion that measures mean absolute error (L1 Loss) [11]. From these task designs, it is noticeable that the reconstruction task would force the encoder network to generate corresponding feature representations which contain enough information about each image. Therefore, it is also possible to adopt the MIM concept to train Transformer backbone efficiently without any labels.

2.3. Content Pair Prediction Methods

In addition to MIM task for the Transformer backbone pretraining, the Image Pair Prediction as one particular pretext task is also applicable on Transformer backbone. In the Natural Language Processing (NLP), Bidirectional Encoder Representations from Transformers (BERT) adopted two pretext tasks in the pretraining stage [12]. Specifically, BERT adopted Masked Language Modeling and classification head for Next Sentence Predication [12]. In reality, BERT utilized Masked Modeling before the developing of ViT, and this masked modeling is also verified by the vision domain as what has been illustrated before. However, the Image Pair Predication, which is similar to the Next Sentence Predication in BERT pretraining tasks, was not studied comprehensively. Therefore, it is worthy to import Image Pair Predication pretext task to assist training of the Transformer backbone model, which could directly help the representation learning.

III. APPROACH

3.1. Training Scheme Design

In this training scheme, the model backbone is still the same compared with vanilla ViT, but the input and output of the model is specialized for the CSL method, which is approximately double the input and output size. Hence, the training scheme is named Double ViT (DViT). Double ViT configured with 12 encoder layers, 4×4 patch size, 3×4×4=48 hidden size, 4×48=192 Multi-Layer Perceptron (MLP) hidden units, 1e-6 epsilon value for layer norm, and 3 heads in Multi Head Self Attention. All of these configurations are down scaled corresponding to the standard ViT-base model. The input is arranged by applying linear projection on two images sequentially, then all image patches would become flattened embeddings. Next, an extra classification token would be appended at the first position of all embeddings, and all embeddings would be summarized with their corresponding position embeddings. Following this, all of them would be fed into the model backbone, then the transformer encoder would perform feature refinement on each embedding. Afterwards, the output embedding corresponding to the classification token would be provided to the classification head for 0 or 1 image pair prediction. Meanwhile, a projector would be applied to all output embeddings corresponding to the first image patches and the second image patches separately, then both feature vectors would be calculated Cosine Embedding Loss after further processing by an extra projection head on each of vector. To

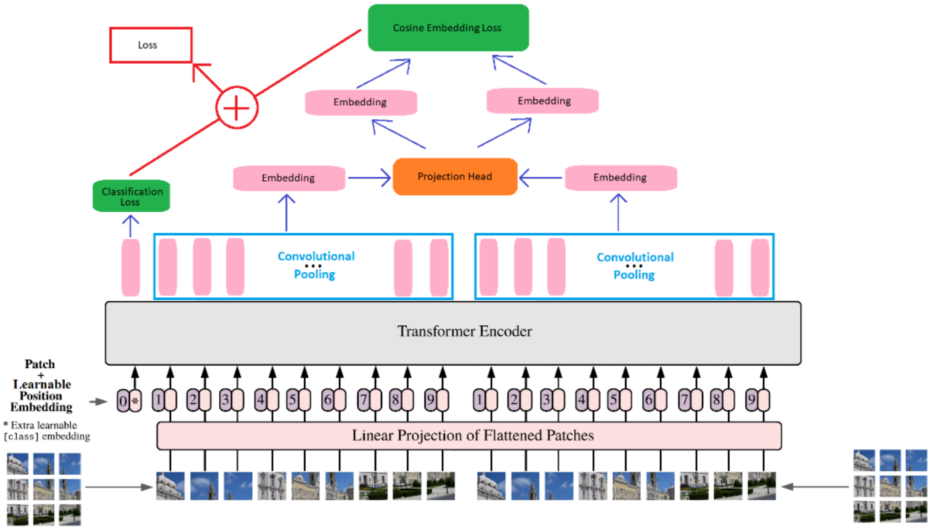


Fig. 1. Pretraining of the Double ViT (Modified from the Model Overview Diagram of ViT [13])

clearly illustrate the proposed training scheme, the model overview was provided as Fig. 1.

The proposed new algorithm adopted two loss function joint training scheme, including the classification loss for the image pair predication, and the contrastive loss for the feature representation vector learning. The classification loss and the contrastive loss are comparable numerically, and the total loss was calculated by summing these two losses up directly. To evaluate the capability of this new algorithm, actual class labels would be involved in the pretraining stage. This is because the algorithm adopted hard contrastive loss to discriminate each sample in the training set. Specifically, the 0 or 1 classification loss does not consider any potential positive image pairs in a batch of training examples, and the adopted CSL loss which is Cosine Embedding Loss also applied -1 or 1 to indicate negative or positive image pairs. Although this two-loss joint scheme could support the model to learn robust representations, it will also generate a gap to common vision tasks such as image classification. Therefore, this new algorithm cannot be expected to achieve relatively better performance on the image classification compared with specialized deep neural networks.

3.2. Settings in Training and Linear Evaluation

All experiments on Double ViT were performed under the same settings except the model backbone and projection head. The pretraining step adopted 45000 images (90% training data) from the CIFAR-10 training set, and the remaining 5000 images (10% training data) would be utilized in the following training for linear evaluation purpose. Specifically, a fully connected layer would be trained with these 5000 images and labels with batch size equals to 512 in 20 epochs on extracted features from the pretrained Double ViT. Meanwhile,

all parameters in the Double ViT model backbone would not be tuned in this process. Only the fully connected layer has changeable parameters in this step, and this training step was depicted as Fig. 2.

The separation of these training images was referenced in the section Image Classification (CIFAR-10) on Kaggle in the book Dive into Deep Learning [14]. To simplify the training

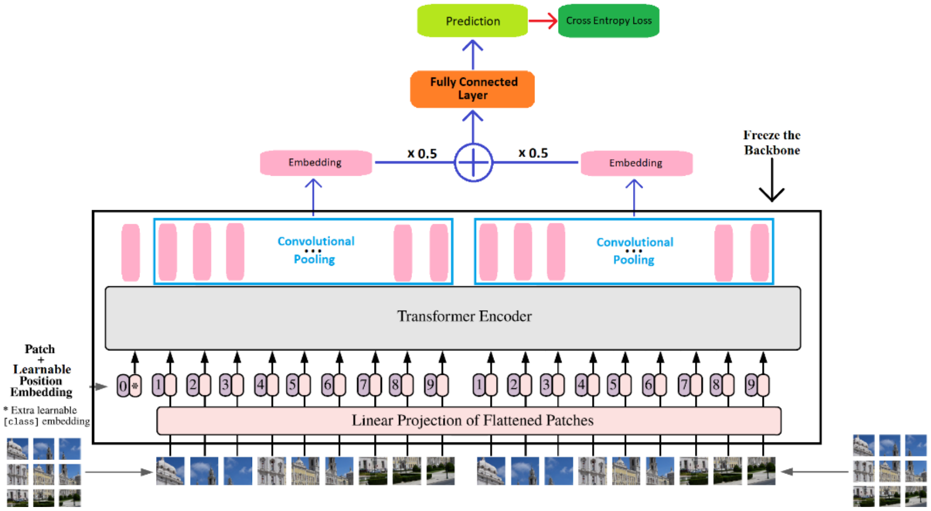


Fig. 2. Training of a Fully Connected Layer for Linear Evaluation

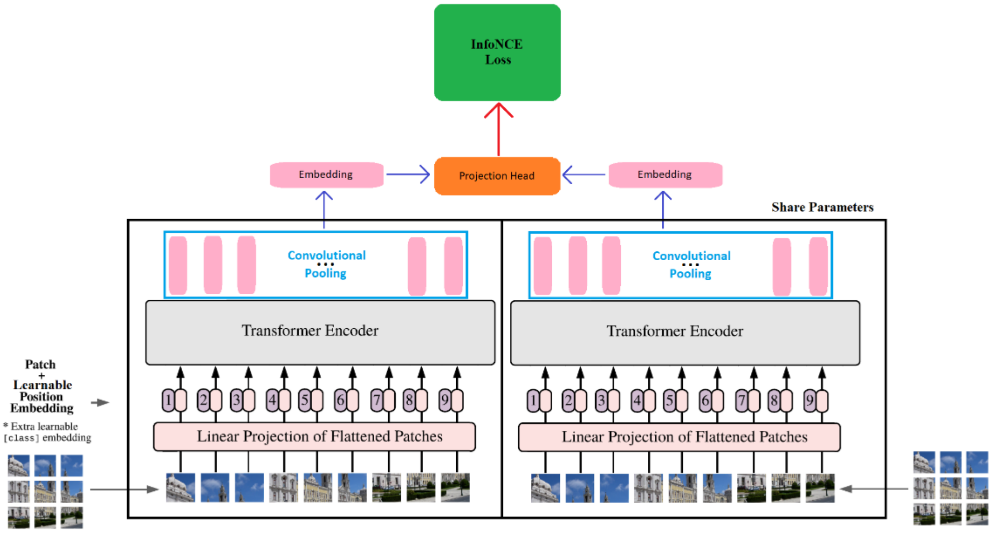


Fig. 3. Pretraining of PoolViT

pipeline, there is no data augmentation involved in the Double ViT models pretraining. The optimizer AdamW was configured with the common learning rate lr=1.25e-3 and weight_decay=1e-6. The learning rate scheduler adopted warm up and cosine decaying tricks, which involves 5 equivalent epochs linear increasing, and 5 equivalent epochs identity, then apply cosine decaying for remaining epochs. At inference time, two same images would be fed into the Double ViT sequentially, then apply the projector on both outputs with averaging each of output embedding to form the feature representation vector. The pretraining epochs of all models is equivalent to 800, because all image pairs selected in training a standard Double ViT is split by class labels, which also requires random selection in 1/10 training samples. Thus, the standard Double ViT needs 8000 iterations to achieve equivalent 800 epochs training result.

IV. CIFAR-10 EXPERIMENTS

4.1. Experimental Environments

All experiments were conducted with Pytorch 1.12.1, Python 3.9.12 and CUDA 10.2. Hardware accelerators are two NVIDIA GeForce RTX 2080 Ti with 11GB dedicated memory on each device.

4.2. Experimental Results and Analysis

To perform a fair comparison among conventional methods and Double ViT training scheme, a model was constructed with the vanilla ViT and a convolutional pooling layer, which is named PoolViT. PoolViT would follow the same training scheme of SimCLR, as demonstrated in Fig. 3.

By comparing different methods fairly, there are three settings are identical in all experiments shown in table 1, which consists of random normal initialization for the learnable position embedding, Embedding Dimension equals 512, and hidden size for output MLP is 2048. All models were pretrained with batch size equal to 2048.

Model	Contrastive Dimension	Projector	Projection Head	Testing Accuracy
DViT		Conv - MLP		44.74%
Pool ViT	128	Conv Pooling	MLP	45.83%
No cls Token DViT	128	Conv Pooling	MLP	**50.67%**
DViT	128	Conv Pooling	MLP	**64.57%**

Table 1. Parameter Settings and Linear Evaluation Results on CIFAR-10 Test Set of Different Models and Training Scheme

These experimental results verified that the projector layer is indispensable for the contrastive embedding training, and the Double ViT is capable of modeling vision signal in images. There is a considerably large testing accuracy difference between PoolViT and Double ViT, which has a 19% performance gap. However, the pretraining conducted on PoolViT did not involve any labels, and the batch size for conventional CSL tasks are relatively larger than 2048. Thus, Double ViT is valuable for the informative embedding generation with a relatively small batch size. It is also verified that classification aided training is considerably helpful for the ViT backbone model pretraining, especially observing the approximately 14% accuracy gap between two bold scores. Therefore, the classification pretext task on class token is still valuable for the Transformer backbone model pretraining, which has been extensively neglected by researchers.

Embed Dim	Contrast Dim	Projector	Projection Head	Hidden Size	Testing Accuracy
3072	512	Flatten	MLP	8192	69.14%
512	128	Conv Pooling	MLP	2048	67.86%
512	128	Conv Pooling	MLP+Batch norm after each linear layer	2048	**71.13%**

Table 2. Parameter Settings and Linear Evaluation Results on CIFAR-10 Test Set of Double ViT Variants

In Table 2, the model backbone is DViT and the method of the learnable position embedding initialization is consistent with BERT, which is the standard deviation equals to 0.02. In addition to the classification task is considerably helpful for the model pretraining, another important finding is that the convolutional pooling applied on the output feature embeddings is powerful for the feature extraction, and batch norm is also helpful for the model pretraining. Although the model with directly flatten projector could outperform convolutional pooling by approximately 2%, it will become computational unaffordable if the image extended to high resolution. For example, if the input image resolution is the same as common ImageNet samples size during model training, which is 224×224, the directly flatten projector would generate a $(Hidden\,Dimension) \times (Sequence\ Length) = (3 \times 16 \times 16) \times (14 \times 14) = 768 \times 196 = 150{,}528$ dimension vector. If the MLP hidden layer size in the following projection head maintain the same convention, which is 4 times the input size, the hidden layer size would become $4 \times 150{,}528 = 602{,}112$. By calculating the total learnable parameters on the projection head only, the result is equal to $150{,}528 \times 602{,}112 + 602{,}112 + 602{,}112 \times 37{,}632 + 37{,}632 = 113{,}294{,}033{,}664$. If assume all parameters quantized to INT8 mode, which only occupy one byte space, it would still require approximately 105.5GB memory space to store it. However, the kernel size adopted in convolutional pooling could be adjusted by measuring computational requirements, which is more flexible than direct flatten projector. Therefore, the convolutional pooling is a good approach to refine features from output embeddings by weighing the balance between computational requirements and model performance.

Experimental results also demonstrated the benefits of BERT initialization method on the position embedding, it can improve the accuracy by 3% compared with random normal initialization. A possible intuition behind this method is that BERT adopted Next Sentence Predication to predict whether two sentences are adjacent to each other in the sequence order, which is a similar pretext task compared with Image Pair Prediction. In both two pretext tasks, the number of inputs is two on semantic level. Thus, it is reasonable to consider that the BERT initialization method could achieve better performance on two inputs training scheme.

V. CONCLUSIONS

In summary, Double ViT model could learn relatively good feature representations by image pair predication task, which is verified by experimental results conducted with linear evaluation on CIFAR-10 dataset. It is noticeable that the projection head in common CSL methods is also efficient for the pretraining of Double ViT to capture vision features in images, and the convolutional pooling is one superior option to down sample output embeddings. Meanwhile, the BERT initialization method adopted on position embedding also presented contribution to the representation learning. Consequently, all experimental results provided expectation that Double ViT is capable of learning high quality representation for images.

VI. FUTURE WORK

A noticeable drawback impedes SSL pretraining of Double ViT is that all class labels were adopted in this training process. One reason for adopting actual class labels in Double ViT is that positive or negative image pairs classification and cosine embedding contrastive are hard contrastive learning tasks. They will not consider any potential positive image pairs in a batch of images, but the temperature parameter τ in InfoNCE loss could count a certain range of positive sample pairs. Thus, it is reasonable to adopt the soft contrastive loss functions such as modified NCE loss to replace the current training scheme, then the real SSL application capability of Double ViT would be extended by excluding the labelling process.

This proposed training scheme demonstrated a powerful learning representation capability on images. However, deep learning models are also capable of capturing patterns for different datasets. For example, it is possible to recognize untrusted environments by performing pairwise comparison at training stage for integrated circuit design. Therefore, it is valuable to conduct the study of trusted environments for integrated circuits design.

VII. ACKNOWLEDGEMENT

This work is supported by the Xi’an Jiaotong-Liverpool University (XJTLU) AI University Research Centre, Jiangsu (Provincial) Data Science and Cognitive Computational Engineering Research Centre at XJTLU; and research funding: XJTLU-REF-21-01-002. The authors do not declare any conflicts of interest.

REFERENCES

[1] W. Falcon and K. Cho, “A Framework For Contrastive Self-Supervised Learning And Designing A New Approach,” arXiv.org, Aug 2020. [Online]. Available: https://arxiv.org/abs/2009.00104

[2] W. Falcon, (2020, Sep. 3). A Framework For Contrastive Self-Supervised Learning And Designing A New Approach [Online]. Available: https://towardsdatascience.com/a-framework-for-contrastive-self-supervised-learning-and-designing-a-new-approach-3caab5d29619

[3] M. Caron et al., “Unsupervised Learning of Visual Features by Contrasting Cluster Assignments,” Conference on Neural Information Processing Systems (NeurIPS 2020)., Jan 2021, [Online]. Available: https://arxiv.org/abs/2006.09882

[4] K. He et al., “Momentum Contrast for Unsupervised Visual Representation Learning,” Conference on Computer Vision and Pattern Recognition (CVPR 2020)., Mar 2020, [Online]. Available: https://arxiv.org/abs/1911.05722

[5] X. Chen et al., “Improved Baselines with Momentum Contrastive Learning,” arXiv.org, Mar 2020. [Online]. Available: https://arxiv.org/abs/2003.04297

[6] X. Chen, S. Xie, and K. He, “An Empirical Study of Training Self-Supervised Vision Transformers,” International Conference on Computer Vision (ICCV 2021)., vol. 33, Aug 2021, pp. 22243-22255. [Online]. Available: https://arxiv.org/abs/2104.02057

[7] J. Grill et al., “Bootstrap Your Own Latent: A New Approach to Self-Supervised Learning” arXiv.org, Sep 2020. [Online]. Available: https://arxiv.org/abs/2006.07733

[8] M. Caron et al., “Emerging Properties in Self-Supervised Vision Transformers,” International Conference on Computer Vision (ICCV 2021)., Oct 2021, pp. 9630-9640. [Online]. Available: https://arxiv.org/abs/2104.14294

[9] H. Bao et al., “BEIT: BERT Pre-Training of Image Transformers,” International Conference on Learning Representations (ICLR 2022)., Jun 2021. [Online]. Available: https://openreview.net/forum?id=p-BhZSz59o4

[10] K. He et al., “Masked Autoencoders Are Scalable Vision Learners,” arXiv.org, Dec 2021. [Online]. Available: https://arxiv.org/abs/2111.06377

[11] Z. Xie et al., “SimMIM: A Simple Framework for Masked Image Modeling,” arXiv.org, Apr 2022. [Online]. Available: https://arxiv.org/abs/2111.09886

[12] J. Devlin et al., “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” North American Chapter of the Association for Computational Linguistics (NAACL 2019)., vol. 1, Jun 2019, pp. 4171-4186. [Online]. Available: https://arxiv.org/abs/1810.04805

[13] A. Dosovitskiy et al., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” International Conference on Learning Representations. Representations. (ICLR 2021)., Vienna., Austria, 2021. [Online]. Available: https://openreview.net/forum?id=YicbFdNTTy

[14] A. Zhang et al., “Image Classification (CIFAR-10) on Kaggle,” in Dive into Deep Learning, 2nd ed. Mar 2022, ch. 13, sec. 13, pp. 646-653. [Online]. Available: https://d2l.ai/chapter_computer-vision/kaggle-cifar10.html

Enterprise-level Corporate Performance Framework for Smart Manufacturing: A Research Framework

Dr Jean-Yves LE CORRE

Abstract— **Significant industry trends have seen companies in China taking advantage of smart manufacturing technologies to lead digital transformation and build competitive advantage on domestic and global markets across several industries. Several survey studies from academic and professional experts have aimed to measure the contribution of smart factories to productivity and business performance in general, however, few studies have investigated the real business impact of Chinese firms' investment in smart manufacturing technologies from a multi-level perspective, by addressing internal and external factors that affect the effect of firm's investment in smart manufacturing technologies towards financial performance. This study aims to propose a research framework for measuring the impact of firms' investment in smart manufacturing technologies at enterprise (corporate) level, through sample selection methods, the evaluation of intangible assets based on the six capital categories defined by the International Integrated Reporting Council, and key ratios of financial performance. This study will contribute to evaluate the performance of firms investing in smart manufacturing from the point of view of external stakeholders as well as for educational purpose to train executives on monitoring performance of those firms.**

Index Terms— **Smart Manufacturing, Corporate Performance, Integrated Reporting, Performance Indicators, Management Accounting Education**

I. INTRODUCTION

The purpose of the study is to investigate a research framework to measure the impact of Research & Development (R&D) investment in smart manufacturing technologies on financial performance; the business research framework should be derived from theoretical models available in existing literature to measure the performance of R&D investment in smart manufacturing technologies at corporate (enterprise) level, rather than develop a theoretical model.

There are many factors that can affect the financial performance of enterprises which invest in smart manufacturing technologies. The influence of those factors on business performance is diverse, therefore the effect of intelligent manufacturing investment to enterprise performance should be evaluated from both tangible benefit and intangible perspective. The main assumption of our study is that intermediary channels that affect enterprise financial performance from R&D investment in smart manufacturing technologies can be classified according to the six capital categories as defined by the International Integrated Reporting Council). The study builds on the approach developed Ye et al. (2020) which used propensity score matching with difference in differences (PSM-DID) method to investigate the impact of intelligent manufacturing on financial performance and innovation performance. A critical review of the article is provided as well as recommendations to apply the model to the current study.

II. LITTERATURE REVIEW

Few studies have focused on measuring corporate performance of intelligent manufacturing investment. This may be due to the lack of data available in the public domain at the enterprise level (Cheng et al. 2019). Most studies have focused on either national or industry-level. Moreover, the types of data needs to be relevant to intelligent manufacturing as defined as industry 4.0 (Kiel et al. 2017; Lin et al. 2018).

Several studies have investigated the mediation channels that affect the financial performance of intelligent manufacturing industries in the Fourth Industrial Revolution. Yang et al. (2020) argued that innovation performance is a major mediating factor for improving the performance of firms in the intelligent manufacturing industry, assuming that is sufficient to promote financial performance.

III. METHODOLOGY

The research method comprise three main steps as indicated in Figure 1. below.

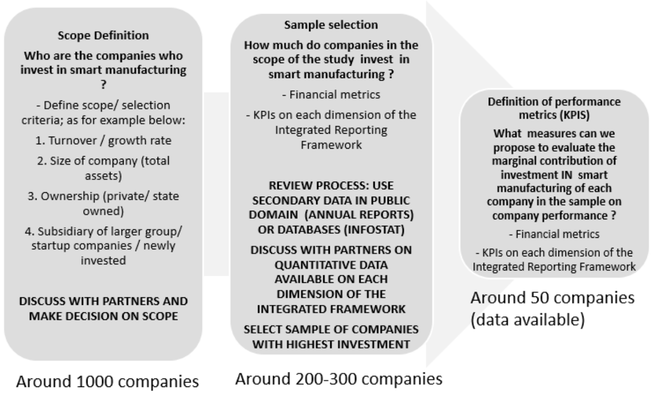

Scope Definition
Who are the companies who invest in smart manufacturing ?
- Define scope/ selection criteria; as for example below:
1. Turnover / growth rate
2. Size of company (total assets)
3. Ownership (private/ state owned)
4. Subsidiary of larger group/ startup companies / newly invested
DISCUSS WITH PARTNERS AND MAKE DECISION ON SCOPE
Around 1000 companies
Sample selection
How much do companies in the scope of the study invest in smart manufacturing ?
- Financial metrics
- KPIs on each dimension of the Integrated Reporting Framework
REVIEW PROCESS: USE SECONDARY DATA IN PUBLIC DOMAIN (ANNUAL REPORTS) OR DATABASES (INFOSTAT)
DISCUSS WITH PARTNERS ON QUANTITATIVE DATA AVAILABLE ON EACH DIMENSION OF THE INTEGRATED FRAMEWORK
SELECT SAMPLE OF COMPANIES WITH HIGHEST INVESTMENT
Around 200-300 companies
Definition of performance metrics (KPIS)
What measures can we propose to evaluate the marginal contribution of investment IN smart manufacturing of each company in the sample on company performance ?
- Financial metrics
- KPIs on each dimension of the Integrated Reporting Framework
Around 50 companies (data available)

Fig. 1. Research Framework - Overview

A. Sample Selection

The public database is used to collect intelligent manufacturing data based on proprietary patent value model, which integrates the most common and important technical indicators in the patent analysis industry (such as technical stability, technological advancement, protection scope, etc. including the type of patent, the number of citations, the number of the same family, the number of countries of the same family, the number of claims, the number of inventors, the number of large groups involved in IPC, the remaining validity period of the patent, etc.), and divides patents into 1-10 points by setting parameters such as index weight and calculation order. Further sampling selection can be made on segments of the top three smart manufacturing companies, or analysis of the composition of patented technologies in the field of smart manufacturing of top three companies can be made directly based on patent search results. To complete the sample selection, it is decided that: (1) When analyzing a single enterprise, the global analysis is based on the country to which the enterprise belongs; the nationwide analysis is based on the province and city where the enterprise's industrial and commercial registration address belongs. (2) For the overall patent analysis, the global layout analysis and the geographical distribution analysis of applicants can be carried out according to the country of patent disclosure and the country of patent application (that is, the country where the patent application is accepted) in the patent search results; It is applicable to global national organizations or regions, and Chinese provinces and cities.

B. Measurement of Intangible Capital

After samples selection, the measurement of environmental, social, economic and governance performance of company is completed through content analysis and text coding to mine the information related to intelligent manufacturing from the financial report of listed enterprises with 'intelligent manufacturing' and related domains as the keyword.

C. Financial Performance

Data acquisition is mainly from financial databases such as the Wind Financial Database. Overall measures can be those related to shareholder return (shareholder value added) or return on investment (e.g. ROI), or value added (e.g. EVA, MVA). Specific measures might be units or value of sales, profit margin, or relationships between those measures through ratios (e.g., productivity).

IV. CONCLUSIONS

A final stage, a composite formula is proposed to attribute a score to each enterprise of the sample under review. The current study will aim to propose such framework by identifying evaluation and selection criteria to come out with coherent framework for measuring the impact of investment in smart manufacturing technologies on corporate performance in Chinese economic and finance context. To do so, the study develops a performance evaluation framework addressing the marginal contribution of smart manufacturing technologies on each component of integrated reporting The framework is further evaluated and assessed through a thorough peer-review processes by a group of academic and professional experts.

REFERENCES

[1] Yu, Yubing & Zhang, Justin & Cao, Yanhong & Kazançoğlu, Yiğit. (2021). Intelligent transformation of the manufacturing industry for Industry 4.0: Seizing financial benefits from the supply chain relationship capital through enterprise green management. Technological Forecasting and Social Change. 172. 120999. 10.1016/j.techfore.2021.120999. RFC6120 Extensible Messaging and Presence Protocol (XMPP): Core. P. Saint-Andre. March 2011

[2] Nguyen, Hanh & Ha, Minh-Tri. (2020). Social capital and firm performance: A study on manufacturing and services firms in Vietnam. Management Science Letters. 10. 2571-2582. 10.5267/j.msl.2020.3.038.

[3] Jie Yang, Limeng Ying & Manru Gao (2020) The influence of intelligent manufacturing on financial performance and innovation performance: the case of China, Enterprise Information Systems, 14:6, 812-832, DOI: 10.1080/17517575.2020.1746407

[4] Cheng, H., R. Jia, D. Li, and H. Li. 2019. "The Rise of Robots in China." Journal of Economic Perspectives 33 (2): 71–88. doi:10.1257/jep.33.2.71.

[5] Kiel, D., J. M. Müller, C. Arnold, and K.-I. Voigt. 2017. "Sustainable Industrial Value Creation: Benefitsand Challenges of Industry 4.0." International Journal of Innovation Management 21 (8): 1740015. doi:10.1142/S13639196174001

Optimizing Small Files Operations in HDFS File Storage Mode

Yi-Yang Chen, Rui-Jun Wang, Zhen Hong, Zahid Akhtar, Kamran Siddique

***Abstract*—Hadoop Distributed File System (HDFS) is based on Google File System (GFS), a big data distributed file management system included in Hadoop. Nowadays, many HDFS and many other similar frameworks have the need to store small files in the system. In this aspect, HDFS affects its performance and Namenode memory management when dealing with a large number of small files. Therefore, researchers have proposed various solutions to address the shortcomings of HDFS for storing small and medium-sized files. This paper presents three HDFS schemes for merging small files and analyzes the importance of correlation and prefetching after merging small files. The efficiency of reading small files can be improved by correlated file prefetching. Finally, the small file storage architecture is obtained to stand superior to the NHAR architecture.**

***Index Terms*— Hadoop, HDFS, Distributed file system, small file.**

I. INTRODUCTION

NameNode and several data nodes comprise Hadoop Distributed File System (HDFS) clusters. Multiple clients access HDFS, as demonstrated in Fig. 2 [2]. Appending to HDFS is limited to a single client. HDFS divides file content into 128MB blocks by default. However, users can change this file by file.

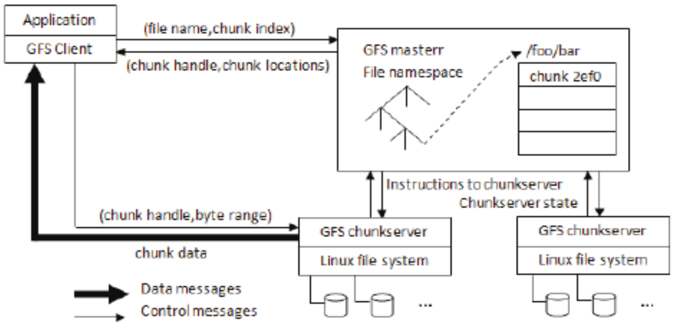


Fig. 1. GFS Architecture [1].

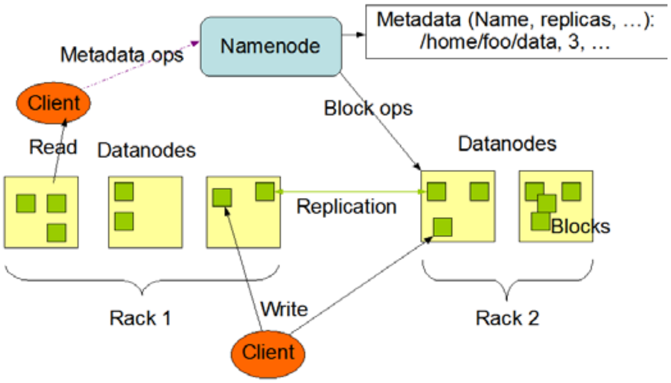


Fig. 2. HDFS Architecture [2].

Table 1. Comparing the differences between GFS and HDFS [3].

	GFS	HDFS
Platform	Linux	Cross Platform
Developer	Google	Yahoo and open source framework
Release Year	2003	2010
Main Server	Master	Name node
Data Server	Chunkserver	Data node
File Server	Chunk	Block
Availability	Operation log	Journal, edit log
Read and write	Multiple writer, multiple reader model	Single writer, multiple reader model
Other operation	Random file writes possible	Only append is possible
Default block size	64 MB	128MB

NameNode is the master-slave architecture process in HDFS that leads the slave nodes, controls the file system namespace, and maintains the file system tree and all files and directories inside it [4].

This information is permanently stored on the local disk in two files: the File System Image (FSImage) and the edit. The FSImage holds information about the HDFS directory tree, meta-information, and data block indexes at a particular moment in time. The subsequent changes to this information are stored in the edit log, which provides a complete NameNode first relationship.

The NameNode keeps track of the DataNode where each block in each file is located. However, the block location information is not stored permanently. The data node rebuilds this information upon the system starting for this reason. The client interacts with the NameNode to obtain the following information before reading and writing file contents to the DataNode. In addition, the NameNode receives information on

Manuscript received October 8, 2022.

[1]School of Computing and Data Science, Xiamen University Malaysia, Sepang 43900, Malaysia
[2]Department of Network and Computer Security, State University of New York Polytechnic Institute, Utica, 13502, USA).

HDFS's overall status, such as the system's available space, the space already used, the current status of each data node, and so on.

Secondary NameNode is a daemon that merges namespace images and image edit logs on a regular basis. To relieve pressure, NameNode stores the files on a disc and passes them to the Secondary NameNode instead of merging and editing the FSImage itself. Each cluster contains a second NameNode, similar to Namenode. In large-scale deployments, the second NameNode usually takes up a single server.

The secondary NameNode differs from the NameNode in that it does not receive or record any real-time changes to HDFS, instead of fetching the FSImage and the edit at a predetermined point [5]. And then merges them to create a new namespace mirror based on the cluster's time interval settings. The secondary NameNode works with the NameNode to provide a simple checkpoint mechanism to prevent the NameNode from taking too long to start due to a big edit log.

Downtime can be decreased, and the danger of NameNode metadata loss can be lowered by checkpointing a second NameNode. However, the second NameNode does not support automatic recovery from NameNode failures. Therefore, NameNode failures must be handled manually.

The file system's working nodes are known as DataNode. They store and retrieve data blocks as needed (subject to the client or NameNode scheduling), and send a list of the blocks they store to the NameNode regularly [5]. Although HDFS is designed for large files, the files stored on HDFS are similar to traditional file systems in that they are also chunked and then stored. However, unlike traditional file systems, on a DataNode, HDFS file blocks (aka data blocks) are stored as normal files on a Linux file system. Slave nodes on an HDFS cluster each reside in a DataNode daemon to perform the busiest part of the distributed file system: writing HDFS data blocks to, or reading data blocks from, actual files on the Linux local file system. Each DataNode informs the NameNode of the currently stored data blocks upon initialization. During their work, the DataNode continues to update the NameNode, supplying it with information about local alterations and receiving requests to create, transfer, or delete data blocks on the local disc from the NameNode.

NameNode employs a file merging approach to decrease the number of files it manages [6]. All files that belong to a large/logical file are merged into a single file, taking file correlations into consideration. Due to HDFS's access method, reading files from it has a high access latency, especially when accessing large numbers of tiny files. Prefetching reduces access latency, but HDFS does not presently support prefetching to disguise I/O latency. If file correlations are disregarded for data placement and prefetching techniques, reading small files from HDFS typically results in a lot of seeking and bouncing from DataNode to DataNode to get files. Therefore, a two-level correlation-based file prefetching strategy that comprises block prefetching and metadata prefetching can be advantageous for Hadoop-based Internet applications. However, the accuracy of file correlation predictions is related to the performance of prefetching associated files. Associated file prefetching leads to the high concurrency of reading requests and prefetch requests when high traffic and accuracy are low. When the prefetched block is stored there, each one needs to be sent to the client. This exacerbates network transfer overload, leading to frequent access to some of the DataNodes. These actions result in a large number of invalid I/O operations and network broadcasts that degrade system performance.

The remainder of the paper's format is as follows: Section 2 contains a literature review, and section 3 analyzes how HDFS merges small files. The results of the analysis are discussed in Section 4, and section 5 provides a conclusion.

II. RELATED WORKED

Previous research has established the detailed principles and features of Hadoop and HDFS, but the researcher discovered an apparent theoretical gap in the earlier research about small file storage. The theory on small file storage is rather dated, and the current studies bear the fruit of this theoretical gap. Hadoop archiving (HAR) is primarily used to archive files in HDFS in order to reduce Namenode memory consumption [7]. Reading small files using HAR is slower than reading them from the original HDFS because HAR requires access to two index files in order to acquire a file [8]. Many researchers have proposed changes to the HAR architecture, such as the New Hadoop Archive (NHAR), to reduce metadata storage requirements and enhance the efficiency of accessing small files [9]. However, Addressing Hadoop's Small File Problem with an Appendable Archive File Format [6], which focuses exclusively on the efficiency of Small File Access. However, an investigation in terms of file correlations and theoretical development is warranted. An investigation of these issues is important because the performance of prefetching correlated files has relations to the accuracy of file correlation predictions [6]. Therefore, the models of this field need to embrace contemporary research in correlation and related fields to provide a stronger theoretical basis for projects [10]. In this article, we compare the previous papers, sort out a new order in the messy discussion, and provide a clear reference for later researchers.

III. METHODOLOGY

The main direction of our research is the small file storage mode of HDFS. We used Data study to collect a large amount of literature and data on optimizing HDFS for small file storage. Then select the three most representative optimization solutions by deeply analyzing the underlying implementation of these three optimization solutions. We identify the differences in the underlying implementation of these three methods. Then we collect information from other papers and study the advantages and disadvantages of these underlying different point processing solutions in depth.

Finally, conclusions are drawn, and suggestions are given for optimizing the HDFS small file storage scheme.

The small file storage model of HDFS is the main focus of our study. There are also many limitations in our research methods. Some articles give the corresponding code, but others do not give the corresponding code for their methods. This makes it impossible for us to perform experiments on our own machines. So our data are taken directly from the articles themselves. This may lead to some data that are not very realistic. We chose only three representative methods this time. So it leads to possible limitations of our methodological analysis.

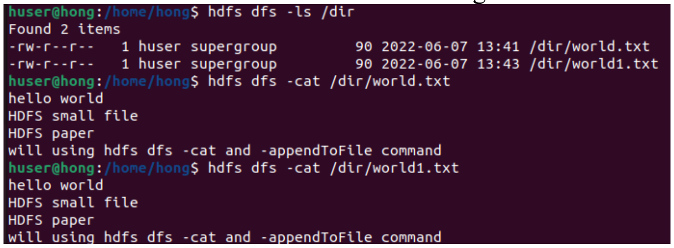
IV. ANALYISIS

A. Motivation

If a file is not 75% of the block size, then it is a small file. The default HDFS block size in Hadoop 3.x is 128MB. Each file stored in Hadoop has at least one block. Each file, directory, and block takes about 150 bytes. If there are too many small files, they can take up a lot of memory. While it is possible to store millions of files on a Namenode, the existing hardware cannot accommodate billions of files [5]. All documents in HDFS are read, and the NameNode controls written processes. Therefore, the greater number of small files, the more frequent the requests to the NameNode [11].

B. Hadoop's own solution

HDFS provides several commands to merge small files for cases where the amount of data is not very large. Under the path of HDFS /dir there are two small files world.txt and world1.txt, the contents of the text file are as follows Fig. 3:

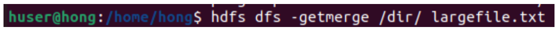


```
huser@hong:/home/hong$ hdfs dfs -ls /dir
Found 2 items
-rw-r--r--   1 huser supergroup         90 2022-06-07 13:41 /dir/world.txt
-rw-r--r--   1 huser supergroup         90 2022-06-07 13:43 /dir/world1.txt
huser@hong:/home/hong$ hdfs dfs -cat /dir/world.txt
hello world
HDFS small file
HDFS paper
will using hdfs dfs -cat and -appendToFile command
huser@hong:/home/hong$ hdfs dfs -cat /dir/world1.txt
hello world
HDFS small file
HDFS paper
will using hdfs dfs -cat and -appendToFile command
```

Fig. 3. Content of each file.

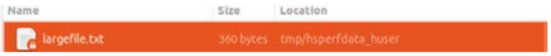
The following two commands can merge HDFS：

1) Combine multiple small files stored on HDFS, download them, and merge them locally to create one large file. Then use the -getmerge command to merge all the small files under the HDFS /dir path into largefile.txt and store them on the local filesystem.



```
huser@hong:/home/hong$ hdfs dfs -getmerge /dir/ largefile.txt
```

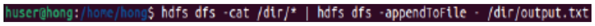
Fig. 4. Command to merge files under the /dir path.



Name	Size	Location
largefile.txt	360 bytes	tmp/hsperfdata_huser

Fig. 5. Command to merge files under the /dir path.

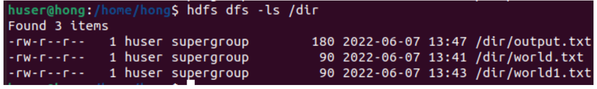
2) Use the command hdfs dfs -cat /dir/* | hdfs dfs -appendToFile - /dir/output.txt to combine two small files from hdfs into the file /dir/output.txt.



```
huser@hong:/home/hong$ hdfs dfs -cat /dir/* | hdfs dfs -appendToFile - /dir/output.txt
```

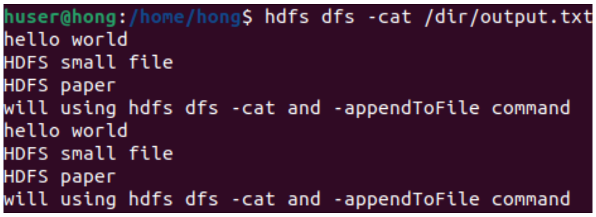
Fig. 6. Use -cat and -appendToFile to merge small files

Then use hdfs dfs -ls /dir to see the /dir path generated a file output.txt, as follows in Fig. 7. Fig. 8 using hdfs dfs - cat /dir/output.txt file merged content.



```
huser@hong:/home/hong$ hdfs dfs -ls /dir
Found 3 items
-rw-r--r--   1 huser supergroup        180 2022-06-07 13:47 /dir/output.txt
-rw-r--r--   1 huser supergroup         90 2022-06-07 13:41 /dir/world.txt
-rw-r--r--   1 huser supergroup         90 2022-06-07 13:43 /dir/world1.txt
```

Fig. 7. List the files under the /dir path.



```
huser@hong:/home/hong$ hdfs dfs -cat /dir/output.txt
hello world
HDFS small file
HDFS paper
will using hdfs dfs -cat and -appendToFile command
hello world
HDFS small file
HDFS paper
will using hdfs dfs -cat and -appendToFile command
```

Fig. 8. List the files under the /dir path

All three of these methods can merge small files. After the small files are merged into a new file, the original file is not deleted file. This processing method is unsuitable for use in the case of huge amounts of da, which will occupy the memory of HDFS. The content of a specific small file cannot be found exactly after the files are merged into one document. We need to use the hdfs dfs -rm root directory to delete the original file. When the amount of data is large, it is recommended to use MapReduce to merge.

C. Small File Storage Architecture

To enable HDFS to support small file processing. The original HDFS storage structure consists of the user layer and storage layer. Some papers propose adding a data processing layer between the user and storage layers.

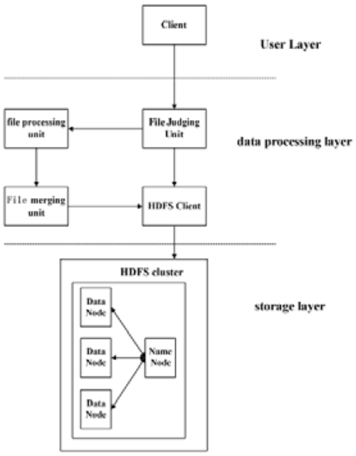


Fig. 9. The storage structure of the improved HDFS system [12].

The data processing layer has four main processing flows. The first is the document judgment unit. This unit is used to determine whether a file is small. If it is a small file, merge it, not store it directly. Next is the file processing unit. This unit counts the size and order of small files and generates temporary index files. The small files are then merged through the file merge unit. Moreover, merge temporary index files to generate merge index files. The merged index files record the small files' index in <key, value> format. The detailed format of the index file is "Hash：<key, offset_length>. The key encodes the small file's important information and is a one-of-a-kind value for retrieving small files. The offset and length of the tiny file in the merged file are recorded in value. As a result, "offset + length" can be used to calculate the small file's end position. In this case, if we want to find a particular small file from the large file that has been merged, we can also find it easily.

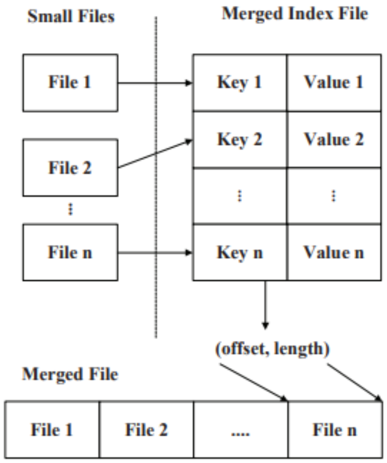


Fig. 10. The index structure of the small file [12]

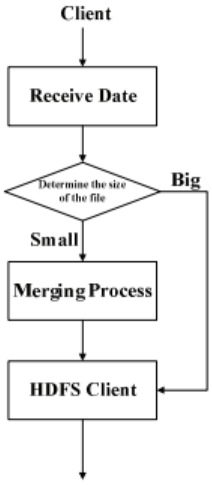


Fig. 11. The processing flow in processing layer [12]

This method makes it easier to find the original small files in the merged large files than manually merging small files.

D. NHAR architecture

To improve access performance, our NHAR architecture adopts a single-level index. Once the time of accessing the index for HAR to read the file is reduced, the access performance can be improved. Instead of using the primary index method, Information about the index is divided into many index files and stored in a hash table by NHAR [9]. It reduces the number of steps necessary to find the file by comparing the hash code to the quantity of index files to find the index file that contains metadata.

For improving efficiency, the process of adding new files to an existing NHAR file using NHAR may be condensed into three steps: archiving the new file, merging the index file, and relocating the new part file. Besides that, the insertion procedure verifies the portion file's overall file size [9]. If the component file's total file size is greater than the maximum block size, the insert procedure will notify the user that the NHAR file needs to be prepared. These portion files are merged to the maximum size of each block after this operation, and the memory usage is lowered [9].

In conclusion, compared with HAR, NHAR's file accessing and adding new files are better; compared with original HDFS, NHAR's archive is much more efficient for storing small files.

V. DISCUSSION

Section IV examined three strategies for merging small files and arrived at the following conclusions. The original HDFS small file merge command merges small files. However, the original files still exist, leading to memory consumption issues. Compared to Hadoop's own solution for accessing small files, small file storage architecture and NHAR architecture are superior for accessing small files. The accessing of small files is more efficient than Hadoop's own approach. Comparing the NHAR architecture with the small file storage architecture, we get the result that the latter solution is superior. The reasons are as follows.

The associated files' metadata and blocks will be prefetched from NameNode and DataNode [11]. The small file storage architecture prefetching scheme effectively prevents NameNodes overload. After the first reading of small files, the merging index file is stored in the cache to speed up the read speed [9]. Prefetching and caching techniques are used to reduce the load on the NameNodes. According to Huang et al. (2018), merging files based on the degree of association between small files will improve the effectiveness of prefetching [13]. The probability of querying a file in the cache will increase in the next access. Prefetching hides and reduces visible I/O costs by leveraging correlation between files [7]. It improves access times and even mitigates network transfer latency by fetching data to the cache before it is requested [8].

Prefetching functionality is not available in HDFS. There is currently just one static read-ahead technique available. After prefetching the entire block asynchronously, one buffer at a time is read synchronously from the block prefetching results, irrespective of the amount of data requested for the block.

Prefetching is believed to reduce access latency effectively. HDFS prefetch is more complex than traditional file system prefetch due to HDFS features. Currently, HDFS does not provide the prefetch function, and this results in a heavy load on Namenode when accessing a large number of small files. Small File Architecture's approach solves this problem.

To control how much data to prefetch and when to trigger the next prefetching request, the prefetch degree and trigger distance are dynamically adjusted [14]. They are the key areas of attention for existing adaptive prefetching algorithms.

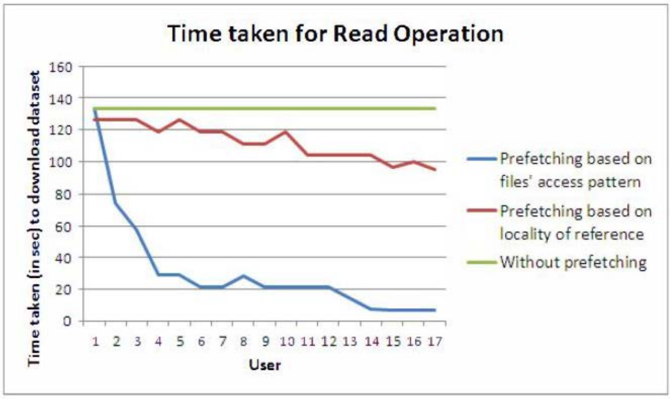


Fig. 12. Time taken for Read operation [15].

Effective consideration is given to file correlations and access locality while storing and retrieving small files. First, to lessen the metadata burden on NameNode, all correlated small files are combined into a single larger file. In order to increase the speed of reading small files, a two-level prefetching mechanism—which includes prefetching of both indexed files and data files—is added. When users access a file, the indexed file in the associated block is added to memory, eliminating the need for NameNode interaction. This process is known as "indexed file prefetching." Data files prefetching entails that after users view a file, the entire block in which the file is located is loaded into memory. If a file is accessed, the Priority Heap of the preceding file is obtained. The current file item is examined in the priority heap. If it does, the file's priority is increased by one. If not, a new file entry with priority 0 is added. Collecting prefetching information from the heap and collecting and caching the relevant files if they are accessible for the current file. So, clearly, speed will increase as the user visits more files. On the web server, files are prefetched into the cache, removing cache coherency problems. Prefetching based on file access patterns speeds up time by 94% compared to not prefetching and cuts read access time by 92% compared to the reference location [15].

This method may result in "hot" DataNodes that regularly serve requests in HDFS prefetching. The HDFS cluster's workload balance must, therefore, also be taken into account [11].

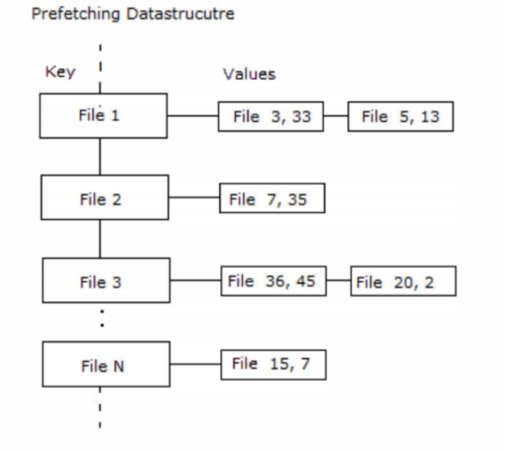


Fig. 13. An Example of Prefetching Date of files [15].

Although NHAR improves access efficiency, there are still many problems. For instance, increased memory consumption in the internal NameNode and creating of NHAR files without considering the correlation between small files. Therefore, the disadvantages of NHAR architecture are more obvious compared to small file storage architecture.

The user launches the create operation by using the client-side module for the Extended Hadoop Distributed File System (EHDFS). The client is then given an output stream and other helper methods that help link file data with a particular element in the combined file. In order to correlate file data with an entry in the combined file based on the data transmitted to the output, the client is then given an output stream and other support techniques. The information output to the output. As we can see in Figure 14:

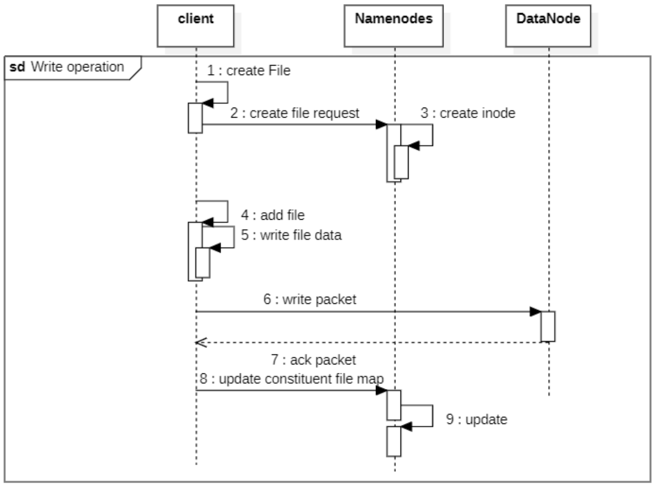


Fig. 14. Write operation in EHDFS [16].

In the read operation, if the metadata already exists in the cache. When a client opens a small file, no request is sent to the NameNode. It will read directly from the cache. As we can see in Figure 15:

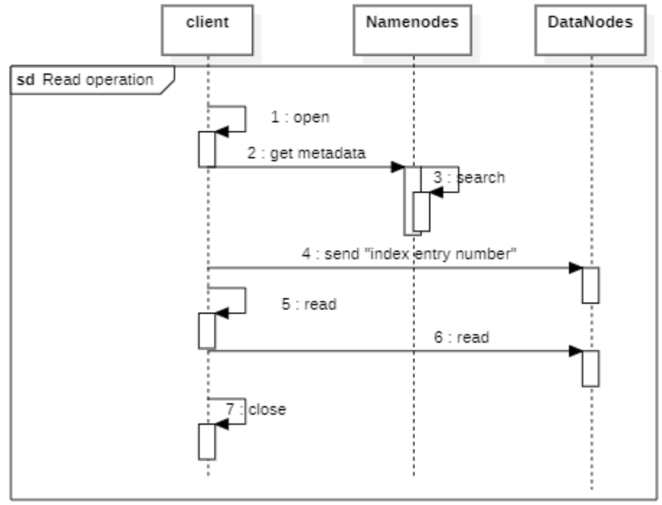


Fig. 15. Read operation in EHDFS [16].

VI. CONCLUSIONS

HDFS is a distributed file management system in Hadoop. It is designed to manage large files. So when HDFS is used for a large number of small files, it results in wasted memory of NameNode and inefficiency of MapReduce. Then, we describe in detail three HDFS methods for merging small files.

First, the HDFS system merges small files command is the basic method which lacks some features such as not being able to batch process small files and small files merge without indexing. Therefore, more methods have been proposed to solve these problems, such as NHAR architecture. This is one of the most valuable approaches because it greatly improves the efficiency of accessing small files. However, this approach does not take into account the correlation between small documents, so there are still some shortcomings that need to be mentioned, such as it has no way to prefetch other small files when reading a small file based on the relevance of the small file. Following these foundations is the small file storage architecture strategy. This method solves the problem of file prefetch and improves the efficiency of reading small files.

The majority of this article's analysis takes the form of a survey, which is devoid of empirical support and must be followed up by future experiment design. The analysis in this paper is subjective, and the accuracy of efficiency assessment among these methods needs to be further improved in the future.

There is no unified solution for merging small files. To address the effects of small files on HDFS performance and NameNode memory management, an efficient small file merging approach should be suggested.

REFERENCES

[1] Ghemawat, S., Gobioff, H. and Leung, S., 2003. The Google file system. ACM SIGOPS Operating Systems Review, 37(5), pp.29-43

[2] HDFS Architecture Guide. Hadoop.apache.org. (2022). Retrieved 21 May 2022, from https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html

[3] Miles, D.A. (2017). A Taxonomy of Research Gaps: Identifying and Defining the Seven Research Gaps, Doctoral Student Workshop: Finding Research Gaps - Research Methods and Strategies, Dallas, Texas, 2017

[4] Dhage, S., Subhash, T., Kotkar, R., Varpe, P., & Pardeshi, S. (2020). An Overview - Google File System (GFS) and Hadoop Distributed File System (HDFS). SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology, 12(SUP 1), 126-128.

[5] Ghemawat, S., Gobioff, H. and Leung, S., 2003. The Google file system. ACM SIGOPS Operating Systems Review, 37(5), pp.29-43

[6] Renner, T., Müller, J., Thamsen, L., & Kao, O. (2017). Addressing Hadoop's Small File Problem With an Appendable Archive File Format. Proceedings of the Computing Frontiers Conference. https://doi.org/10.1145/3075564.3078888

[7] HDFS Architecture Guide. Hadoop.apache.org. (2022). Retrieved 21 May 2022, from https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html

[8] Patel, A., & Mehta, M. A. (2015). A novel approach for efficient handling of small files in HDFS. 2015 IEEE International Advance Computing Conference (IACC). https://doi.org/10.1109/iadcc.2015.7154903

[9] Vorapongkitipun, C., & Nupairoj, N. (2014). Improving performance of small-file accessing in Hadoop. 2014 11th International Joint Conference on Computer Science and Software Engineering (JCSSE). https://doi.org/10.1109/jcsse.2014.6841867

[10] Miles, D. (2017), A Taxonomy of Research Gaps: Identifying and Defining the Seven Research Gaps

[11] Dong, B., Zhong, X., Zheng, Q., Jian, L., Liu, J., Qiu, J., & Li, Y. (2010). Correlation Based File Prefetching Approach for Hadoop. 2010 IEEE Second International Conference on Cloud Computing Technology and Science. https://doi.org/10.1109/cloudcom.2010.60

[12] Changtong, L. (2016). An improved HDFS for small file. 2016 18th International Conference on Advanced Communication Technology (ICACT). https://doi.org/10.1109/icact.2016.7423437

[13] Huang, L., Liu, J., & Meng, W. (2018). A Review of Various Optimization Schemes of Small Files Storage on Hadoop. 2018 37th Chinese Control Conference (CCC). https://doi.org/10.23919/chicc.2018.8483588

[14] Dong, B., Zheng, Q., Tian, F., Chao, K. M., Ma, R., & Anane, R. (2012). An optimized approach for storing and accessing small files on cloud storage. Journal of Network and Computer Applications, 35(6), 1847–1862. https://doi.org/10.1016/j.jnca.2012.07.009

[15] Aishwarya K, Arvind Ram A, Sreevatson M C, Babu, C., & Prabavathy B. (2013). Efficient prefetching technique for storage of heterogeneous small files in Hadoop Distributed File System Federation. 2013 Fifth International Conference on Advanced Computing (ICoAC). https://doi.org/10.1109/icoac.2013.6922006

[16] Liu, X., Peng, C., & Yu, Z. (2015). Research on the Small Files Problem of Hadoop. Advances in Social Science, Education and Humanities Research. https://doi.org/10.2991/emcs-15.2015.9

Assessment of Organ Equivalent Dose & Effectual Dose from Diagnostic X – Ray in Gombe Specialist Hospital: A Case Study

Muhammad Mudassir Usman, Abdullahi Muhammad**,** Nuruddeen Muhammad Abdulkareem & Kabiru Hamza

mudather1197@gmail.com

***Abstract*--The radiation dose that the patient receives during the radiological examination is crucial for preventing exposure concerns; research is being done to acquire effective doses for common diagnostic radiographic examinations at specialist hospitals. Gombe, the entrance surface dose and effectual dose received by patients undergoing diagnostic X-ray examinations were directly estimated by the research. Measuring parameters included x-ray dose output, backscatter factor, focus to skin distance, as well as physical parameters like and in mathematical model. The study's findings indicated that for the chest (PA, AP), abdomen (AP), and cranium (AP, lateral), the mean entrance surface doses and effectual doses obtained are 0.2432mGy, 0.2857mGy, 0.6331mGy, 0.7553mGy, 0.3220mGy, 0.0121mGy, 0.0142mSv, 0.07597mSv, 0.00755mSv and 0.00322mSv, respectively. It was found that the doses received by patient in Gombe Specialist Hospital does not exceed the international standard as given by WHO.**

***Index Terms*- Radiation, Exposure, Equivalent, Examination**

I. INTRODUCTION

X-rays are electromagnetic radiation that selectively penetrates body structures and produces images of those structures on fluorescent or photographic film [3]. Diagnostic X-rays are the images in question. Diagnostic X-rays are helpful in identifying bodily anomalies. They provide a quick, painless technique to rule out issues like broken bones, tumors, tooth rot, and the presence of foreign bodies. X-rays easily flow through air and soft bodily tissue, but they are halted when they come into contact with more dense objects like tumors, bones, or metal fragments [8].

Positioning the bodily part to be studied between a focused X-ray beam and a plate containing film is how diagnostic X-rays are carried out. This process doesn't hurt. The amount of x-ray absorption increases with the density of the substance that the x-ray passes through [7]. As a result, bone absorbs more radiation than muscle or fat, and tumors may do the same relative to the surrounding tissue. A portable x-ray machine interacts with the x-ray that travels through the body and hits the photographic plate.

Today, diagnostic x-ray exams are a common and significant source of all medical exposures employed in global population's medical diagnosis [5]. The use of this has increased due to the high demand for X-ray exams in developing nations and the rise in X-ray machine numbers. Therefore, it is necessary to enforce radiation protection while projecting radiography images in accordance with the concepts of rationale, optimization, and individual dosage limits in order to reduce the associated dangers.

A doctor's and physicist's assurance that the doses received by radiographic patients should be in accordance with the principle of As Low as Reasonably Achievable (ALARA) and that the dose does not exceed the amount necessary to get proper radiographic imaging is made possible by quality control and dose measurement [9]. Dosage assessment is essential to improve the optimization of the radiation protection of the patients and to provide the lowest dose possible during exams in the field of radiology because the ionizing nature of X-rays means that their usage is not entirely risk-free [11].

[9] evaluated the entrance skin doses for 500 patients having six different types of diagnostic X-ray examinations, including the entrance skin doses. Inferred from measurements and understanding of X-ray output variables was the entrance skin. We employed questionnaire physical characteristics like mAs and kV along with measurements parameters like X-ray dosage output, back scatter factor, and focus to skin distance. For entrance skin dosages to the chest PA, skull AP, abdomen, c-spine, pelvic AP, hand, and foot, the means and standard deviations were, accordingly.

II. METHODOLOGY

Twenty adults, mixed-gender patients were engaged on this research, which was conducted at specialist hospital Gombe, Gombe state, test for the chest, cranium, and abdomen in radiography department, patient exposure parameters were collected. Anterior Posterior and Posterior Anterior (PA) chest, Anterior Posterior (AP) abdomen, and Posterior Anterior (PA) and Lateral (LAT) head are the six most often conducted diagnostic x-ray examinations that are included in the sampling. Tube potential (kVp), exposure setting or current (mAs), and film focus distance (FFD) were the radiographic or exposure factors that were typically employed in the radiology room by the radiographer for average-sized adult patients. The user must provide a measured or calculated free-in-air entrance surface dose (ESD in rad), as well as the examination's technique parameters, in order to assess the organ doses from radiography procedures.

The collected data were numerically analyzed, and the following major formulas were applied for the computation based on the goal and objectives of this study. The kerma was obtained in mR directly using the solid-state detector (Radiation meter), and it was then converted to mGy using 1 = 0.876 in accordance with the International Commission on Radiation Units and Measurements.

On this research, the ESD was determined by factoring in X-ray dose output, back scatter factor, focus to skin distance, and physical characteristics like mAs and kV. Once the focus to skin distance (FSD), exposure period, and tube potential and current are determined, entrance surface dosage is stated as (Haval Y Yacoob and Hariwan A. Mohammed).

$$ESD = BSF \times Tube\ output \left(\frac{mGy}{mAs}\right) \times \left[\frac{100}{FSD}\right]^{2} \times mAs \quad (1.1)$$

Table 1. Summary of Patients Characteristics and Technical Parameters Selected for the Various Examination Gombe Specialist Hospital Consider for the Research

S/N	Examination Projection	Sex	Age	Total patients	Tube Potential (kVp)	Exposure Settings (mAs)	Focus to skin Distance
1	PA Chest	M	30	1	65	30	72
2	PA Chest	F	25	1	75	30	63
3	PA Chest	M	20	1	80	32	80
4	PA Chest	F	18	1	83	33	72
5	PA Chest	M	65	1	83	34	72
6	AP Chest	F	50	1	83	36	72
7	AP Chest	M	25	1	84	33	72
8	AP Chest	F	40	1	84	34	87
9	AP Chest	M	50	1	84	35	72
10	AP Chest	F	80	1	85	35	72
11	AP Abdomen	M	25	1	75	40	56
12	AP Abdomen	F	20	1	80	40	56
13	AP Abdomen	M	45	1	85	45	63
14	AP Abdomen	F	30	1	90	50	56
15	AP Abdomen	M	70	1	95	55	63
16	Lat. Skull	F	30	1	65	30	62
17	Lat. Skull	M	25	1	80	40	62
18	AP Skull	F	75	1	90	50	62
19	AP Skull	M	45	1	90	55	62
20	AP Skull	F	80	1	95	55	62

Where FSD is the focus to skin distance in centimeters, mAs is the current in milli-amperes, and the backscatter factor, the tube's output is measured in mGy/mAs. At 80 kV, 1 m away, and 10 mAs, the tube calibration is carried out. After that, a radiation weighting factor is added to the entrance surface dose in mGy to get the corresponding dose in mSv (Haval Y Yacoob and Hariwan A. Mohammed).

$$H_T = \sum_T W_T\, D_T \quad (1.2)$$

Effective dosage is defined as the sum of the weighing factor times the equivalent dose in mSv (Haval Y Yacoob and Hariwan A. Mohammed).

$$E = \sum_T W_T\, H_T \quad (1.3)$$

Table 2. Estimated Entrance Surface Dose (ESD) and Effectual Dose for all Projections

S/N	Examination Projection	Entrance surface Dose (mGy)	Equivalent Dose (mSv)	Effectual dose (mSv)
1	PA Chest	0.158115	0.158115	0.007906
2	PA Chest	0.274951	0.274951	0.013748
3	PA Chest	0.206939	0.206939	0.010347
4	PA Chest	0.283594	0.283594	0.014180
5	PA Chest	0.292187	0.292187	0.014609
6	AP Chest	0.309375	0.309375	0.015469
7	AP Chest	0.290468	0.290468	0.015423
8	AP Chest	0.204972	0.204971	0.010249
9	AP Chest	0.308073	0.308073	0.015404
10	AP Chest	0.315451	0.315451	0.015773
11	AP Abdomen	0.463978	0.463978	0.055677
12	AP Abdomen	0.527904	0.527904	0.063348
13	AP Abdomen	0.529737	0.529737	0.063568
14	AP Abdomen	0.835161	0.835161	0.100219
15	AP Abdomen	0.808761	0.808761	0.097051
16	Lat. Skull	0.213234	0.213234	0.002132
17	Lat. Skull	0.430673	0.430673	0.004307
18	AP Skull	0.681339	0.681339	0.006813
19	AP Skull	0.749472	0. 749472	0.007495
20	AP Skull	0.835061	0.835062	0.008351

Table 3. Estimated Effectual Dose Compared With W.H.O Standard

S/N	Examination	Current study (mSv)	W.H.O Standards
1	Chest	0.01216	0.1
2	Abdomen	0.07597	0.128
3	Skull	0.00755	0.01

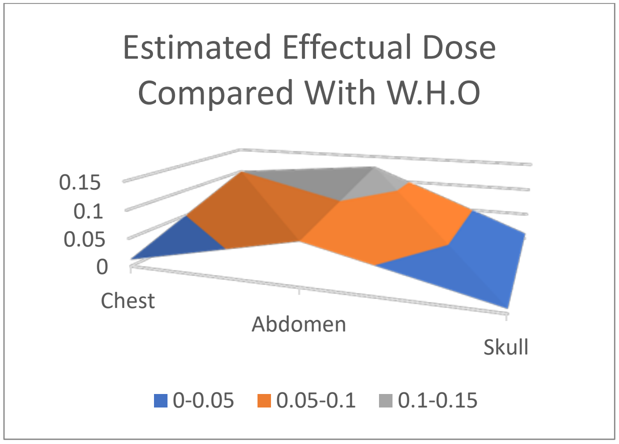


Fig.1. Estimated Effectual Dose Compared With W.H.O

III. Discussion

The European Recommendations on Quality Criteria for Diagnostic Radiography Pictures include diagnostic reference dose levels as one of their quality criteria.

Chest x-ray

The exposure data from the patients were collected throughout the X-ray examinations. Chest (AP) ranges from 0.20497 mGy to 0.315451 mGy and 0.010249 mSv to 0.015773 mSv with mean values of 0.2857 mGy and

0.01428 mSv respectively. The entrance surface dose of the chest (PA) ranges from mGy to mGy and mSv to mSv. Due to the use of low tube current, the ESD (mGy) for the chest (PA) was lower than that published by the IAEA (IAEA, 2005) and European Committee [1]. Compared to stated values in the literature, the effective dose was lower [4].

Abdomen x-ray

Five (5) patients who had conventional abdominal X-rays had their entrance surface doses and effective doses assessed (AP). With mean values of 0.6331 mGy and 0.07597 mSv, respectively, the findings obtained range from 0.463978 mGy to 0.835161 mGy and 0.055677 mSv to 0.100219 mSv. The thickness of the subject being studied may be the reason why the ESD for the AP abdomen was lower than the value published by the European Committee [2].

Skull X-ray

The entrance surface dose and effective dose measured during the X-ray examination of the skull (AP) vary from 0.681339 mGy to 0.83506 mGy and 0.681339 mSv to less than the 0.21234 mGy to 0.430673 mGy and 0.002132 mSv to 0.004307 mSv with mean values of 0.3220 mGy and 0.00322 mSv, respectively. Compared to the figure provided by the IAEA [5] and European Committee, the entrance surface dose (ESD) for the skull (AP) is significantly lower [2]. While the effective dose (ED) of the skull (Lateral) is less than the value published by [1], the effective dose (ED) of the skull (AP) is around the same [1].

IV. CONCLUSION

The research focuses on analyzing the organ equivalent and effective doses from diagnostic x-rays in Gombe specialist hospital, Gombe State, Nigeria. It has been found Patients in specialist hospital Gombe do not receive more indirect ESD than what is recommended by international organizations (WHO and IAEA). Based on the findings of this study, it is possible to draw the conclusion that the use of the right radiological parameters, such as a large patient-to-X-ray source distance, a high tube potential, and a low tube current, can significantly reduce the absorbed dose. This research also demonstrates that when technical and clinical factors are optimized or used properly, patient doses will be significantly decreased.

REFERENCES

[1] Akbar, A., Ehsan, M., Mahboubeh, M., Morteza, S., Mehran, M., (2015). *Measurement of entrance skin dose for common diagnostic X-ray examinations* in Kashan, Iran. Glob J Health Sci. 7(5): 202-207.

[2] European Commission. (2008). European Guidance on Estimating Population Doses from Medical X-ray procedures. Radiation Protection n.154

[3] Darrell R. Fisher and Frederic H. Fahey. Appropriate *use of effective dose in radiation protection and Risk Assessment,* HealthPhys.2017 August; 113(2): 102–109., doi:10.1097/HP.0000000000000674.

[4] *Rahman MS, Sujimura N, Yoshida TY (2008) Characterization of calibration x-ray fields in establishing operational quantities for the AIST middle Beam spectrum series. Bangladesh J Phys 5(6): 39-45.*

[5] *Rasuli B, Mahmoud-Pashazadeh A, Ghorbani M, Juybari RT, Naserpour M (2016) Patient dose measurement in common medical X-ray examinations in Iran. J Appl Clin Med Phys 17(1): 374-386.*

[6] Taha, M.T., Al-Ghorable, F.H., Kutbi, R.A., Saib, W.K., (2015). Assessment of entrance skin doses for patients undergoing diagnostic X-ray examinations in King Abdullah Medical City, Makkah, and KSA. Journal of Radiation Research and Applied Sciences volume 8.

[7] ICRP Publication 103. Ann ICRP. 37(2-4):1-332. http://dx.doi.org/S0146- 6453(07)00039-5.

[8] International Atomic Energy Agency (2007). Technical Reports Series No. 457 Dosimetry in Diagnostic Radiology: An International Code of Practice. Vienna.

[9] International Electro technical Commission, (2002). Medical electrical equipment part 2-44: particular requirement for the safety of X-ray equipment for computed tomography rep IEC6060. ICRP-103., (2007). The 2007 Recommendation of the International Commission on Radiological Protection.

[10] WHO. Quality Assurance in Diagnostic Radiology. Geneva, Switzerland: World Health Organization; 2001.

[11] National Council on Radiation Protection and Measurements. (1990). Implementation of the principle of as low as reasonably achievable (ALARA) for medical and dental personnel: recommendations. NCRP report 107. Bethesda, Md.

Customer Behavioural Trends in Online Grocery Shopping during COVID-19

Kiran Barbole, Ou Liu*

Abstract— ***Online shopping trends have increased over the past few years drastically. The main objective of this research project is to understand the online grocery shopping workflows from the customer's perspective and to identify the increasing trends or patterns observed during COVID-19 in this sector. For the sole purpose of this research project, with the help of the sampling method, approximately 28 respondents were surveyed out of 50 forms sent, and data collection was done through a structured questionnaire. The data analysis was performed based on categorizing the data into 2 main components using the IBM SPSS statistics 28.0.1.1 version analysis tool. We found that consumers were definitely inclined towards OGS services during COVID-19 due to safety, convenience and restrictions imposed by the government. Consumers satisfaction depended on the safety precautions during COVID-19, the assistance provided through helplines for support and increased customer reach, to make the groceries accessible like any other well-known online sector.***

Index Terms— **Customer Behaviour, Online Grocery Shopping, COVID-19**

I. INTRODUCTION

The COVID-19 pandemic caused major outbreak of many large businesses all over the world, out of which especially the Food industry was heavily impacted with negative effects. In the UK, the highest record made in the online shopping sector was around 87% which was 10% higher than 2018 [1]. The start of 2020 introduced a series of events that largely affected the online sector and majority turned out tough on the daily essentials like groceries as there were lockdowns imposed on almost every city in the UK. There were different types of lockdown laws imposed by the UK government in response to the coronavirus pandemic, like gathering, movement and business restrictions [7]. Any kind of contact with a crowd of people was something everyone was scared of. So everyone had to stay home and avoid outside contact. There was a change in behaviour from normal to panic buying/ impulsive buying [2] during the early stages of COVID-19, this change was a motivation in the first place for this research. OGS services is not a new thing, but due to pandemic, there was a whole new group of people who were inclined towards it, and to reduce the gap between this new relationship, the OGS companies took new steps to reach their consumers and provide help/support 24/7 [12].

The main underlying objective of this research is to study the changing landscape and understand the influential factors in the behaviour of online grocery shopping in the midst of pandemic, due to growing concern for cashless shopping options for groceries and daily items. Main research points are mentioned below:

a. To perform an evaluation of demonstrable method to inspect and learn about the majority factors responsible in the customer behaviour.

b. To examine and verify the consumer perception from the data collected for buying groceries online during COVID-19

c. To clearly provide a heuristic approach of the positive and negative experiences of consumers while online grocery shopping, like for those who prefer it and those who don't?

The aim and objective lay the foundation of this research project and after carefully understanding them both, below are the main questions/ problems that this research paper will focus on. The below questions also outlines the tasks that define this project. This research has been planned out, clearly thought of by authors curiosity and motivation, which when combined together would produce valuable and robust findings [3].

a. What are the main influential factors responsible for a consumer choosing the online grocery shopping during COVID-19?

b. What are the essential do's and don'ts for the online grocery shopping industry to take care of during pandemic?

c. How does the consumer experience during the pandemic lay down the future for the online grocery shopping industry as a whole? Gaining advantage from customer learning behaviours

II. LITERATURE REVIEW

Manuscript received Sept. 12, 2022.

Both authors are with Operations and Information Management Department, Aston University, Birmingham, UK (email: 210231861@aston.ac.uk; o.liu@aston.ac.uk).

The emergence of the pandemic introduced many changes in our day-to-day lives, from daily routine to daily shopping. As much as government imposed measures were important, it was also established that individually self-imposed safety measure like frequent hand-washing, wearing masks, sanitizer usage regularly and keeping away from crowds was proved the be the most effective during COVID-19 [11]. There are many existing theories of protection motivation, which arises out of fear and anxiety of an individual and then they begin to take preventive measures to avoid a much more dangerous and extreme tough situation. FEAR is described as a motivational challenging variable for an individual which makes them take necessary action to feel safe and decreases the intensity of a far worse situation [9]. COVID-19 was highly effective situation in bringing the most efficient motivational factor which is fear, and people started taking actions to stay safe and away from the pandemic. [10].

Following the conclusions identified based on the above literature review studies, we want to validate whether the influential factors are positively responsible for the change in consumer behaviour and the new pattern of OGS services. Thus, the below hypothesis have been perceived to analyse the final results of this research :

H1a: AGE is a significant factor in relationship between purchase intentions and online platform

H1b: Profession is a significant factor in moderating the online purchase intentions

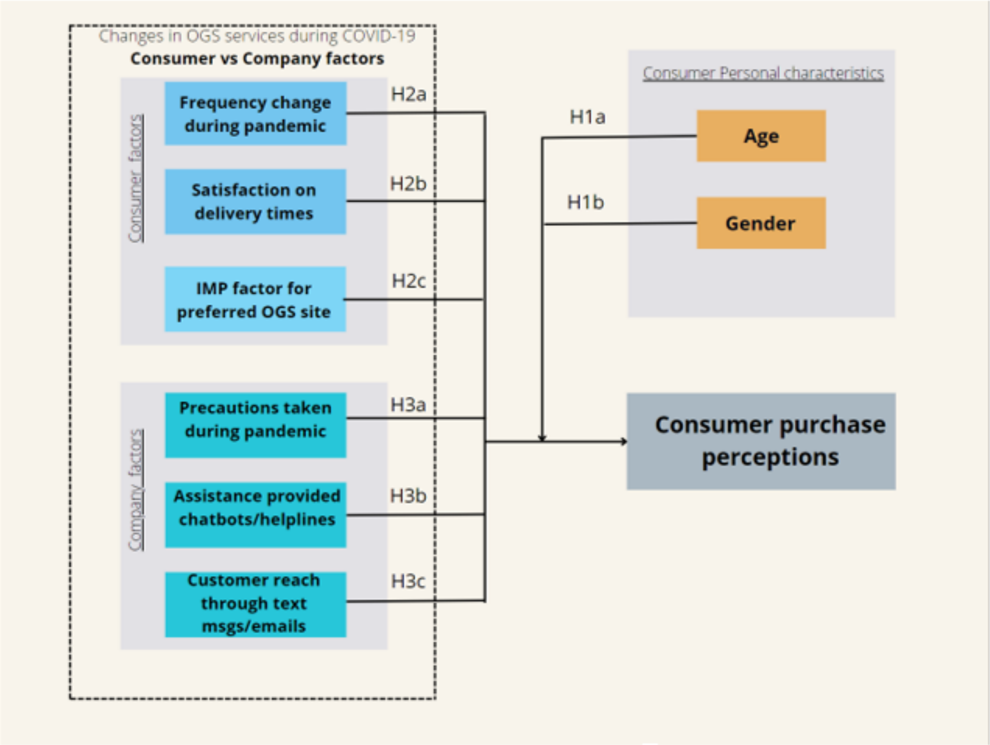

Changes in OGS services during COVID-19
Consumer vs Company factors
Consumer factors
Frequency change during pandemic
Satisfaction on delivery times
IMP factor for preferred OGS site
Company factors
Precautions taken during pandemic
Assistance provided chatbots/helplines
Customer reach through text msgs/emails
H2a
H2b
H2c
H3a
H3b
H3c
H1a
H1b
Consumer Personal characteristics
Age
Gender
Consumer purchase perceptions

FIG 1: *Combining existing and influential factors for changes in consumer behaviour during COVID-19 – Motivational and Perceptual framework*

H2a: Frequency of shopping groceries online have increased during and post-pandemic.

H2b: Customer satisfaction was derived from factors like safety measures, convenience and flexible delivery times and reason for change in consumer perception

H2c: Safety issues, convenience, discounts, quality of products and value provided were significant factors for consumers having a preferred website.

H3a: The precautionary measures taken by the OGS companies, were a significant part of the new approach and reason for change in consumer perception

H3b: Assistance provided through helplines/support pages or emails by the OGS companies was considered by the consumers to feel valued in the OGS services.

H3c: Customer reach through emails, messages and recommendations through strong customer relationship management has a strong effect on customers intention to shop groceries online again.

III. RESEARCH METHODOLOGY

In order to test the hypothesis described and analysed in the literature review, the online questionnaire method was used. A primary advantage of using online questionnaires is that they can save time, cost effective and give easy access to desired community with the same characteristics. Researchers can save a lot of time by working on other tasks and simultaneously collecting the data in the backend [4]. First 5 questions were related to understand the consumers personal OGS characteristics, like starting from their age, profession and "whether they shop for groceries online or no", if they selected yes, then they would be required to jump to the next questions and if they chose NO, then they should select an option in the further question, which clearly specifies the reason for it.

Next section of questions from Q6 – Q11, were to understand their perspective on the OGS services during COVID-19 and what they like/dislike in general. Then, after Q8, the questions were intended towards understanding what motivated them to order groceries from the same OGS website.Q12 onwards it was basically to analyse the behaviour from the OGS companies during COVID-19 to increase sales and provide safe deliveries to consumers. For e.g., whether they followed the safe precautions and rules mentioned by the government.

Data collection was done using the google forms questionnaire and a total of 30 responses were recorded out of 50 forms sent. The survey consisted of 17 questions, which started from the consumers profession, to the experiences they had during pandemic In OGS services, what they liked and what they disliked, how the companies responded to the COVID-19 rules like safety precautions, safe deliveries, etc. The questionnaire was sent to 50 participants living in the UK who were using the OGS services and had access to mobile, laptop, desktop with an internet connection and out of which we successfully got responses from 30 people.

IV. RESULTS

The results of this survey were split into 3 main categories on which the below descriptions and analysis has been derived. The main intention of forming these categories of responses was to analyse the consumer + OGS companies behaviour during COVID-19 and derive the most promising influential factors:

4.1 Data Analysis and Interpretation

Results: Consumer behavioural changes during COVID-19 – Perceptions were divided into 3 sections as below:

1. **<u>Personal factors aligned:</u>**

H1a – Age

H1b – Profession

2. **<u>Consumer factors:</u>**

H2a - Frequency change during pandemic in using OGS

H2b - Satisfaction on delivery times

H2c – Imp factor for preferred OGS website

3. **<u>Company factors:</u>**

H3a – Precautions taken during pandemic

H3b – Assistance provided using chatbots/helplines

H3c – Customer reach through messages/emails

This analysis has been divided into Consumer changes and correspondingly Companies response to COVID-19 restrictions.

4.2 Influential factors in consumer behavioural changes during COVID-19

The analysis of the dependent variable, which is "Frequency of using OGS change during COVID-19" was calculated with the other factors to identify the statistical significant aspect of the hypothesis and understand whether they can be accepted or not.

Q1) Did Frequency of using OGS change during COVID-19? * Were you satisfied with delivery times followed during COVID-19? - The frequency of change in shopping for groceries online during COVID-19 has changed drastically with many new users moving to the remote working and responding proactively to this digital shift. For determining the statistical significance of satisfaction with delivery times on frequency of using OGS during COVID-19, the Pearson's chi-square tests were run and the chi-square statistic was 25.229, the p-value (.014) is in the same row in "Asymptomatic significance (2-sided)" column, which is less than the designated alpha level (.05), hence the result is significant and the researcher rejects the null hypothesis and data confirms that the second variable has a significant influence on the first one and therefore **H2b** is accepted fully.

Q2) Did Frequency of using OGS change during COVID-19? * Important factor in ordering from the same OGS website - To identify the value of customer satisfaction with the regularly used websites for online grocery shopping, the chi-square test result for this relation was found to be 13.755 with a p-value of .317, as the result was more than the normally accepted p-value of .05, the **H2c** hypothesis confirmed that the 2 variables were independent of each other and did not have any statistical significant effect in the outcome. Hence, the **H2c**

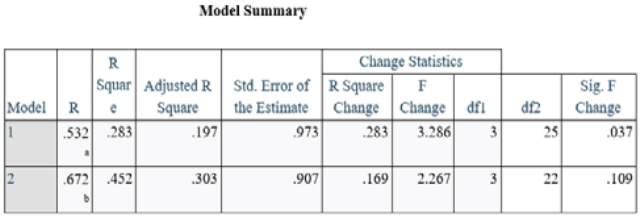
hypothesis was rejected fully as the 2 variables were not significantly related to each other.

Q3) Did Frequency of using OGS change during COVID-19? * Did OGS provide smart assistant s/chatbots? - The third combination was between the frequency of change variable and smart assistants/chatbots, which was statistically proven to be significant with a score of p-value=.053 which has a difference of 0.003 with the normally accepted p-value. The chi-square value was 7.683 with 75% of expected count with less than 5. This data analysis is supposed to be partially accepted for **H3b** hypothesis, as there is slight significance of Smart assistants/ chatbots provided variable on the frequency of change.

Model Significance:

The linear regression model was prepared using the SPSS tool and used to analyse whether the influential factors in different formats add any statistical predictive power to the changes in consumer behaviour/ perception intentions during COVID-19. In the below analysis, there are 2 models specified to analyse the constant and variable factors, like in Model 1, Frequency of change, customer satisfaction and consumer preferred OGS site have found to be statistically significant (F(3,25)= 3.286, p>0.000) with an adj R2 of .197. As the R2 was found to be .283, and the adj. R2 was .197, it indicated a slight reduction in the models precision of statistical significance between the model factors. Thus, Model 1 explains 19.7% of variance in the predicted value of influential factors. It was derived that for identifying the Frequency in change of OGS during COVID-19, there were 2 qualifying metrics like customer satisfaction and delivery times. But due to the reduction in precision, it wasn't correctly defined, which factor had the most weightage.

Statistically, customer satisfaction was found to be significant with (t(1,91)=2.234, p>.000), whereas delivery times was not precisely significant with (t(1,91)=1.445,p=.161). See table 3 for information derived.



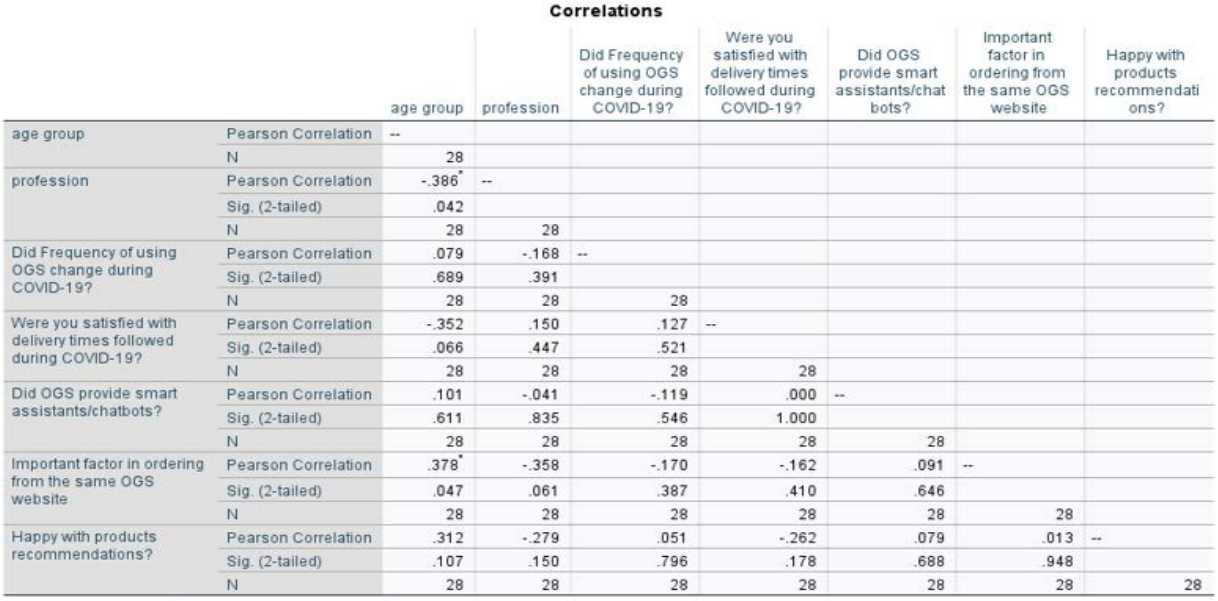
Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.532[a]	.283	.197	.973	.283	3.286	3	25	.037
2	.672[b]	.452	.303	.907	.169	2.267	3	22	.109

FIG 2: Linear Regression Output

With the next 2 influential factors: Precautions taken by the OGS and Assistance provided in the forms of helplines/ support, (F(6,22)= 3.027, p>.000) with an adj. R2 of .303. Hence, the Model 2 explains 30.3% of variance in the predicted values. The frequency of change in online grocery shopping variable was found to be statistically significant with a p value of (p=1.086) in the descriptive analysis.

A linear regression was carried out to identify whether the dependent variable had any statistical significance with the other influential variables that supported the frequency of online grocery shopping during COVID-19. The objective of the Pearson's correlation test was to determine whether the personal factors, in this case: Age



Correlations

		age group	profession	Did Frequency of using OGS change during COVID-19?	Were you satisfied with delivery times followed during COVID-19?	Did OGS provide smart assistants/chat bots?	Important factor in ordering from the same OGS website	Happy with products recommendations?
age group	Pearson Correlation	--						
	N	28						
profession	Pearson Correlation	-.386*	--					
	Sig. (2-tailed)	.042						
	N	28	28					
Did Frequency of using OGS change during COVID-19?	Pearson Correlation	.079	-.168	--				
	Sig. (2-tailed)	.689	.391					
	N	28	28	28				
Were you satisfied with delivery times followed during COVID-19?	Pearson Correlation	-.352	.150	.127	--			
	Sig. (2-tailed)	.066	.447	.521				
	N	28	28	28	28			
Did OGS provide smart assistants/chatbots?	Pearson Correlation	.101	-.041	-.119	.000	--		
	Sig. (2-tailed)	.611	.835	.546	1.000			
	N	28	28	28	28	28		
Important factor in ordering from the same OGS website	Pearson Correlation	.378*	-.358	-.170	-.162	.091	--	
	Sig. (2-tailed)	.047	.061	.387	.410	.646		
	N	28	28	28	28	28	28	
Happy with products recommendations?	Pearson Correlation	.312	-.279	.051	-.262	.079	.013	--
	Sig. (2-tailed)	.107	.150	.796	.178	.688	.948	
	N	28	28	28	28	28	28	28

*. Correlation is significant at the 0.05 level (2-tailed).

***FIG 3**: Pearson's independent variables correlation matrix*

and Profession influenced the change in online shopping behaviour during COVID-19. Also, the other specified hypothesis that have been observed in this matrix, where we tried to inspect the difference in the variables significance in the regression models to analyse the multicollinearity between the influential factors and the frequency for change in using OGS services during COVID-19. As we can see in the below table, the Frequency of change variable has a significant correlation between many other variables. e.g. Frequency of change variable was significantly related with the OGS providing smart assistants/ chatbots (.101), consumers happy with product recommendations (.312) and the important factor that consumers preferred website based on their choices (.378). Hence, the tests analysis and significant variables stated to accept the **H2a** and **H3c** hypothesis completely. This 2-way correlation was identified to be significant at 0.05 level (2-tailed) matrix. The difference between the significant values of the variables was due to multicollinearity.

When adding the 3 influential factors: consumers happy with product recommendations, consumers satisfied with delivery times, Safety precautions taken during pandemic in Model 2 it was found to be statistically significant (F(6,22)=3.027) and p>0.000 with an adj. R2 of .303. After the statistical significance was established between the variables, the points increased by .146 points for each unit in the influential factor of variable customer reach, .240 points for each unit of increase in the consumers satisfied with delivery times during covid-19 variable and safety precautions had an effect with 2.961 points for each unit of increase. Thus accepting the hypothesis, **H1a,H2a,H2b,H3a, H3b, and H3c**.

In this case, due to multiple differences between the profession factors and important factor for consumer preferred OGS site, which is a dependent entity on multiple factors like situational change, product demand, OGS site offerings, etc., the **H1b and H2c** hypothesis were rejected, as the variables were independent and did not have any statistical significance on the relationship between the influential factors for increasing the frequency of consumers using the OGS services during COVID-19.

V. DISCUSSION

The below table describes a quick overview of the accepted/ rejected hypothesis and a summary of the researchers results based on the data analysis and interpretation. Please refer to FIG 06 below, which acts as a guidance. We hypnotized the influential factors for consumers to shop online during COVID-19, the frequency in their shopping and the new pattern emerged from that situation.

One of the most important factors was the safety precautions undertaken by the OGS companies during COVID-19 to assist their consumers. The impact of price value, attitude, and subjective norms has been positive on the online grocery purchase intentions of consumers during COVID-19 [6]. The increase in frequency of using OGS services during COVID-19 was definitely due to the imposed restrictions and consumers had to use the click-and-collect or home-delivery services. This study demonstrated that the influential factors like, safety precautions taken during pandemic by the OGS websites, delivery times followed by them, customer reach through various methods opted and the preferred site chosen by the consumers due to various factors was interrelated and had a significantly strong impact on the increase of frequency for using OGS services. The coefficient matrix was calculated to understand the concept of multicollinearity, whose unwanted relation between independent variables cause the standard error of the coefficients to increase and can eventually cause a statistically significant factor to become insignificant [8]. There were different types of studies/analysis done after the pandemic, which explained how the online grocery shopping transformed and achieved the change of a decade in just few days. McKinsey released several reports regarding the adoption of digital, consumer behaviour changes and consumer satisfaction depending on the new experiences. This research contributes to the existing studies of McKinsey [5], which is similar to the existing research by them, the consumers look for convenience, but in line with assortment, quality and price points, this is value proposition aspect, but the marketing strategies - messages, emails (customer reach) of online grocers should reinforce these elements as well. Our statistical analysis addressed the new pattern of consumer behavioural changes during COVID-19 in the online grocery shopping sector. We identified the insights, the new trends, customer preferences, value-driven behaviours and contribution of OGS services in achieving the new normal.

The Empirical findings suggested that the frequency of using OGS services was increased during pandemic due to the safety measures and risk of contamination, consumers were switching to online services and preferred good quality products, precautionary measures, flexible delivery times and customer satisfaction was directly related to the completion of the above listed points when majority of consumers were transitioning to the digital shift. COVID-

19 introduced a variety of new consumer behaviours in the online grocery shopping sector, as this was a forced behaviour, the guarantee of retaining it is very thin, on the other hand, consumers will stick to it depending on how satisfied they were with different aspects of the services.

VI. CONCLUSION

This study intended to understand and analyse the consumer behavioural trends of online grocery shopping during COVID-19. During this research, we came across many other existing articles and studies that highlighted different aspects of this behaviour. The backbone of this research were the main research questions and the hypothesis conducted out of it, which were a major motivation of the author in conducting the survey and analysing the collected data for effective outcomes. The hypothesis that were developed on the explained literature review and contemporary factors were either accepted or rejected and briefly described in the results section.

We found out that the factors: consumer satisfaction with delivery times and assistance provided by the OGS added statistical power to the existing model in all the regression models analysed. Controversially, the frequency of change in using OGS services during COVID-19 was positively revealed to be significant with the above 2 factors. On the other hand, profession and customer preferences for preferred OGS site was not fully supported and that's why these 2 factors were not found to be relevant. On a whole, the frequency of using OGS services was increased during COVID-19 with supported factors like customer satisfaction with the services provided, concerned safety measures, and assistance from the OGS services was found to be positively influencing consumer behavioural trends in the online grocery context.

We confirmed the statistical significant factors of the conducted hypothesis to satisfy the underlying research, however we encourage other researchers to validate and verify other mediating factors that may be identical/ more beneficial to understand the consumer behaviour. For e.g., the OGS services companies to analyse the consumer shopping patterns during and post-pandemic, smooth digital acceleration in the technologies used with ease and understandable interfaces, the new normal is the new future of the online shopping industry, and this new future depends on the customer satisfaction and their value-driven behaviours. Because the survey was carried out through google forms and distributed over Facebook, Instagram, LinkedIn, and the researchers connections, majority of the participants were students and 78.57% were aged between 20-30 years. Hence, future researchers are encouraged to carry out a probability sampling method to ensure more coverage and broader categories to be considered for a generalized outcome of the study.

REFERENCES

[1] Aashind, A., 2022. 38+ Imposing UK Online Shopping Statistics [2022]. [online] CyberCrew. Available at: <https://cybercrew.uk/blog/uk-online-shopping-statistics/#:~:text=In%202020%2C%2087%25%20of%20UK,the%20UK%20was%20only%2053%25.>

[2] Aljanabi, A., 2021. The impact of economic policy uncertainty, news framing and information overload on panic buying behavior in the time of COVID-19: a conceptual exploration. International Journal of Emerging Markets, [online] Available at: <https://www.emerald.com/insight/content/doi/10.1108/IJOEM-10-2020-1181/full/html>

[3] White, P., 2017. Developing research questions. 2nd ed. Bloomsbury Publishing.

[4] Kevin B. Wright, Researching Internet-Based Populations: Advantages and Disadvantages of Online Survey Research, Online Questionnaire Authoring Software Packages, and Web Survey Services, *Journal of Computer-Mediated Communication*, Volume 10, Issue 3, 1 April 2005, JCMC1034, https://doi.org/10.1111/j.1083-6101.2005.tb00259.x

[5] Galante, N., Monroe, S. and López, E., 2013. The future of online grocery in Europe. [ebook] Available at: <https://www.mckinsey.com/~/media/McKinsey/Industries/Retail/Our%20Insights/The%20future%20of%20online%20grocery%20in%20Europe/The_future_of_online_grocery.ashx>

[6] Tyrväinen, O. and Karjaluoto, H., 2022. Online grocery shopping before and during the COVID-19 pandemic: A meta-analytical review. Telematics and Informatics, 71, p.101839.

[7] Ferguson, D. and Brown, J., 2022. UK Parliament. [online] House of Commons Library. Available at: <https://commonslibrary.parliament.uk/research-briefings/cbp-8875/>

[8] Daoud, J., 2017 Multicollinearity and Regression Analysis. Journal of Physics: Conference Series, 949, p.012009

[9] Ronald W. Rogers (1975) A Protection Motivation Theory of Fear Appeals and Attitude Change1, The

Journal of Psychology, 91:1, 93-114, DOI: 10.1080/00223980.1975.9915803

[10] Eger, L., Komárková, L., Egerová, D. and Mičík, M., 2021. *The effect of COVID-19 on consumer shopping behaviour: Generational cohort perspective*. [online] ScienceDirect. Available at: <https://www.sciencedirect.com/science/article/pii/S0969698921001089>

[11] Teslya, A., Pham, T., Godijk, N., Kretzschmar, M., Bootsma, M. and Rozhnova, G., 2020. Impact of self-imposed prevention measures and short-term government-imposed social distancing on mitigating and delaying a COVID-19 epidemic: A modelling study. *PLOS Medicine*, 17(7), p.e1003166.

[12] Mathew, A., 2022. How Live Chat can Enhance Customer Service for Online Grocery Stores. [online] LiveAdmins. Available at: <https://www.liveadmins.com/blog/how-live-chat-can-enhance-customer-service-for-online-grocery-stores/>

Traffic Accident Scene Recognition with FMCW Radar and Vision Transformer

Runwei Guan[1,2,3], Ka Lok Man[1], Liye Jia[1,2], Yuanyuan Zhang[1,2], Shanliang Yao[1], Eng Gee Lim[1],
Jeremy Smith[3] and Yutao Yue[2]

Abstract— **Deep learning has been widely used to classify and detect the FMCW radar targets of traffic vehicles. However, the recognition work of traffic accidents based on the millimeter-wave radar is much rare and complicated. Besides, constructing complex target datasets such as traffic accidents is expensive and laborious. Therefore, this paper proposes an emulation framework to generate maps of some traffic accident scenes and generate targets of traffic accidents in range-azimuth (RA) maps and elevation-azimuth (EA) maps based on FMCW radar. In addition, deep learning with self-attention is applied to classify traffic accidents in EA and RA maps while detecting in RA maps. The results show that the vision-transformer-based neural network performs better than the CNN-based neural network in recognizing complex traffic targets. The main code of the project is available at https://github.com/GuanRunwei/mmWave-Accident-Scene-Emulation.**

Index Terms— **FMCW radar; emulation; traffic accident; vision transformer; self-attention; convolutional neural network.**

I. INTRODUCTION

As a significant perception sensor in autonomous vehicle (AV) [1] system and advanced driver assistance system (ADAS) [2], millimeter-wave radar has the advantages of low cost, strong anti-interference ability and all-weather work [3]. Millimeter-wave radar can detect range [4], doppler [5], azimuth [6], radar cross section (RCS) [7] and phase information [8] of targets, so as to perform target detection on the radar's range-azimuth, range-doppler and range-azimuth-doppler maps. In addition, with the advent of 4D radar, millimeter-wave radar can measure the elevation information of the target and get radar images with richer point cloud [9].

With the gradual maturity of deep learning applied to radar target recognition, there have been many successful application cases of deep neural network (DNN) [10], which includes target classification and detection for radar.

Cai et al. [11] proposed 4 classification models with CNN based on different types of radar data including RCS, range-azimuth maps and 3D radar point cloud images. With the development of vision transformer, Bai et al. [12] proposed a radar data backbone based on self-attention architecture to classify target data collected by TI MIMO radar.

The above works fully show the excellent work on deep learning in the territory of radar target recognition. However, there is little work on identifying traffic accidents or events based on FMCW radar. As a kind of event, the traffic accident is also an essential component of the traffic, but the data are hard and dangerous to collect. Therefore, this paper proposes an emulation architecture based on Radar Toolbox in MATLAB to generate the different traffic accidents and detect targets by an emulational FMCW radar. In addition, we use vision-transformer-based [13] neural networks to classify the radar targets in RA and EA maps while detecting in RA maps, which includes individual traffic vehicles and complex scenes like traffic accidents. At the same time, CNN-based neural networks are also used in the experiments, whose results are compared with those of vision-transformer-based neural networks.

II. RELATED WORK

A. Fundamentals of Millimeter-wave Radar

FMCW radar is a form of CW radar where the carrier frequency used to modulate the signal is linearly varied with respected to time. The FMCW signal sent from the transmitter (Tx) is composed by a number of unity chirp signals generated periodically. The receiver (Rx) captures the echo signal who has the altered frequency and phase. This section will describe the principle of extracting the range and angle information of the targets based on the FMCW radar system.

For range estimation, the transmitted signal sent from the FMCW radar can be expressed as

$$x_{Tx}(t) = A_{Tx}cos\left(2\pi f_c t + \pi \frac{B}{T_c} t^2\right) \tag{1}$$

where A_{Tx} is the amplitude of the transmitted signal, f_c is the starting frequency of the chirp signal, T_c and B are the period and bandwidth of one chirp signal, respectively. The phase information is neglected in range estimation because it can be removed easily by the local oscillator (LCO) according to the range-correlation effect [14]. Once encountering the target in a distance of R with a speed of v, the radar signal will be reflected as

$$x_{Rx}(t) = A_{Rx}cos\left(2\pi f_c t + \pi \frac{B}{T_c} (t - t_d)^2\right) \tag{2}$$

where A_{Rx} is the amplitude of the received signal and $t_d = \frac{2(R+vt)}{c}$ is the round-trip time delay with light speed c. The received signal will be first mixed with the transmitted signal

[1]School of Advanced Technology, Xi'an Jiaotong Liverpool University, Suzhou,China({Runwei.Guan21,yuanyuan.zhang16,Liye.Jia17,shanliang.yao19}@student.xjtlu.edu.cn; {Ka.Man, enggee.lim}@xjtlu.edu.cn)

[2]Institute of Deep Perception Technology, JITRI, Wuxi, China ({guanrunwei, yueyutao, zhangyuanyuan, jialiye}@idpt.org)

[3]Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, UK({runwei.guan, J.S.Smith}@liverpool.ac.uk)

as $x_m(t) = x_{Tx}(t) \cdot x_{Rx}(t)$ and then passed through the low-pass filter to obtain the intermediate frequency (IF) signal as

$$x_{IF}(t) = A_{IF} cos\left(2\pi f_c t_d + 2\pi \frac{B}{T_c} t_d - \pi \frac{B}{T_c} t_d^2\right) \quad (3)$$

where A_{IF} is the amplitude of the IF signal. By assuming that T_c and is small enough and the high-order term are negligible, $x_{IF}(t)$ can be further expressed as

$$x_{IF}(t) = A_{IF} cos\left(2\pi\left(f_c \frac{2BR}{T_c c} t + \frac{2R}{c}\right)\right) \quad (4)$$

Therefore, the intermedia frequency can be obtained by simply FFT and the range information is estimated as

$$f_{IF} = \frac{2RB}{T_c c} \; and \; R = \frac{f_{IF} T_c c}{2B} \quad (5)$$

For angle of arrival (AoA) estimation, to estimate the AoA in terms of elevation θ and azimuth ϕ, single-input multiple-output (SIMO) antenna system is adopted in this paper. Different from the 1D receiver array, here the receiver array has a dimension of $N \times N$ and the distance between the adjacent elements is d. The combined IF signal at the receiver array is

$$x(\theta, \phi) = \sum_{k}^{N} \sum_{l}^{N} S_{k,l} e^{j\frac{2\pi c}{f_c}[(k-1)\,\mathrm{d}\sin(\theta)\cos(\phi) + (l-1)\,\mathrm{d}\cos(\theta)]} \quad (6)$$

where $S_{k,l}$ is the signal received from the receiver element index by (k, l). By simply calculating the ratio of the power reflected along the elevation/azimuth coordinate to the overall power across each coordinate, AOA of the targets can be estimated by selecting the area with the maximum power ratio.

B. 2D Object Classification and Detection

Image classification based on end-to-end neural networks such as multi-layer perceptron (MLP), convolutional neural network (CNN), and vision transformer has proven powerful capabilities. CNN has been developing very fast over the past few decades, including ResNet [15] and VGGNet [16], etc. Vision-transformer-based neural network like ViT conducts image classification by discarding the inductive bias and learning it by the network itself.

2D object detection can be mainly divided into one-stage and two-stage target detection. The one-stage target detection is represented by the YOLO [17] series. The two-stage object detection is represented by Faster R-CNN [18], which consists of a convolutional neural network and a region proposal network. With the development of vision transformer, object detection frameworks based on vision transformer architecture like DETR [19] have been proposed.

III. EMULATION FRAMEWORK OF TRAFFIC ACCIDENT

The emulation framework mainly consists of three parts, which are a target model generator, a millimeter-wave radar emulator and a radar map generator. The entire framework architecture is shown in Fig.1.

A. Target Model Generator

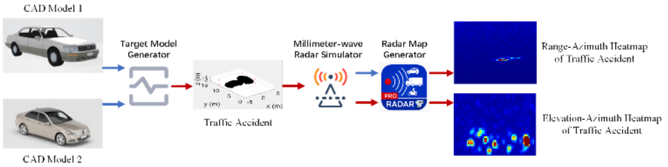

CAD Model 1
CAD Model 2
Target Model Generator
Traffic Accident
Millimeter-wave Radar Simulator
Radar Map Generator
PRO
RADAR
Range-Azimuth Heatmap of Traffic Accident
Elevation-Azimuth Heatmap of Traffic Accident

Fig. 1. The emulation framework of traffic targets

The target model generator takes charge of generating the traffic accident scenes. CAD models are loaded into the generator to generate traffic objects with different poses. Fig.2 shows the examples of generated scenes.

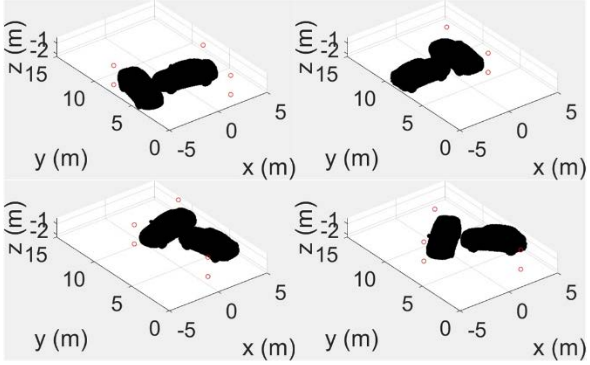


Fig. 2. Traffic accidents with different poses generated by the target model generator.

For the traffic accident scenes, we set a total of 5 typical traffic accident scenes on roads according to the common sense, which are the most common in daily traffic, including the collisions of the two motor vehicles, one motor vehicle and one non-motor vehicle, two non-motor vehicles, one motor vehicle and one pedestrian, and one non-motor vehicle and one pedestrian. After the CAD models required in different scenes are loaded into the generator, the models are arranged and combined to the suitable distances and poses by the corresponding rotation and translation matrices to generate the simple traffic accident scenes. Finally, we randomly treat the generated traffic accident scene as a whole target and generate them of different ranges and azimuths.

Due to the limitation of the maximum angular field of view, sometimes radar could see the part of the target no matter for a single traffic object or traffic accident scene. However, we still include these kinds of data as part of the dataset for training and test, which could enhance the performance and robustness of the model.

B. Millimeter-wave Radar Simulator

Millimeter-wave radar simulator simulates a millimeter-wave radar in *MATLAB* by setting the parameters and formula calculation of millimeter-wave radar. We simulated a 77GHz mmWave antenna array with an angular field of view of 120°, a pitch angle of 60°, an angular resolution of 0.125°, a sweep time of 0.8ms and a frequency sample of 512. There are 1 transmitters (TX) and 40 receivers (RX). It could detect the azimuth, range and elevation of the target by transmitting and receiving the radar signal and process them with fast Fourier transform (FFT). The millimeter-wave radar simulator is placed in a 3D space with coordinates (0m,0m,2m), facing the vehicle targets being scanned. The detected objects and scenes are distributed between 3m and 20m from the radar and their azimuth may exceed the radar's angular field of view. Fig.3 shows the range-azimuth locations of the radar and the targets. Table 1 shows some significant parameters of the radar.

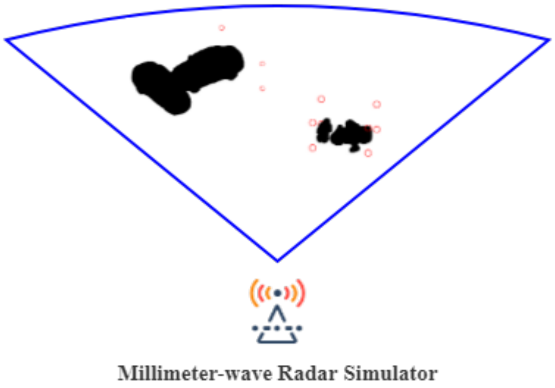


Fig. 3. Range-azimuth Location of Radar and Targets

Table 1. Significant Parameters of the FMCW Radar

PARAMETER	VALUE
Starting Frequency	75.8GHz
Bandwidth	1.2GHz
Sweep Time	0.8ms
Baseband Sampling Rate	500kHz
TX Antenna Position	44cm to the right of RX
Antenna Element Spacing	0.25 cm
TX Number	1
RX Number	40

C. *Radar Map Generator*

The radar map generator is used to generate range-azimuth maps and elevation-azimuth maps of the detected target by FFT. The two maps with bins in azimuth and elevation dimensions are transformed into corresponding heatmaps according to the signal intensity. We set blue as the base color and the area with a denser target signal has a warmer color. At the same time, we also consider clutter that we add random Gaussian noise to the background.

IV. RADAR TARGET RECOGNITION

The recognition work can be divided into 2 parts, which are classification and detection respectively. The classification work is conducted in both EA maps and RA maps while the detection work is only conducted in RA maps.

A. *Data Augmentation*

Data Augmentation plays an essential role in image classification and object detection, which could avoid over-fitting while enhancing robustness and generalization. Firstly, the images are all resized to the same size, after that there are 3 kinds of augmentation methods used in the dataset, which respectively consider the background color, image blur and image rotation as shown in Table 1. The augmentation samples are shown in Fig.4.

Table 1. Data Augmentation of Image Classification

BackgroundColor	Blur	Rotation
ColorJitter	MotionBlur	RandomRotation
RandomGrayscale	MedianBlur	ShiftScaleRotate

B. *Classification*

There are 4 neural network models considered in the classification part, which are ResNet, EfficientNet, ViT and TNT. The first two are typical convolutional neural networks while the last two are typical vision transformer. The purpose that we use vision-transformer-based neural network is to identify whether the self-attention mechanism could improve the classification accuracy of the complex scenes such as different traffic accidents.

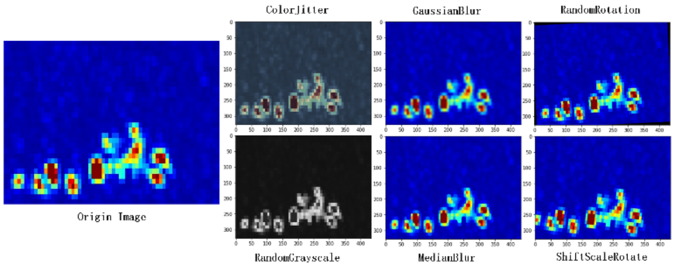


Fig. 3. Data Augmentation (an EA map sample of two non-motors collision)

The classification data consists of 5 kinds of traffic accident scenes, which include two-vehicle collision, vehicle-non-motor collision, vehicle-pedestrian collision, two-non-motor-vehicle collision, and non-motor-pedestrian collision. For each accident scene, there is a RA map and an EA map respectively.

For targets in EA maps, the shape and resolution of the targets are close to optical images. To enhance the model performance and speed up convergence, all the models are loaded with the parameters which are trained by *ImageNet* [20] dataset.

For targets in RA maps, the target shape and resolution have large difference between the images in *ImageNet* dataset, so we did not use pretraining.

No matter for the models used for classification on EA maps or RA maps, fine-tuning is used that the weights of the shallow network are freezed and allow the weights of the deep part of the network could participate in backpropagation to update.

For pretrained models, the output size of the last fully connected layer is adjusted to the number of classes, which is 5 in our experiments. The output is calculated by *softmax* function, which is shown in Eq.7. x_i refers to the current class while x_j represents all classes.

$$Softmax(x_i) = \frac{\exp(x_i)}{\sum_{\mathrm{j}}^{\mathrm{N}} \exp(x_j)} \tag{7}$$

Due to *softmax* would result in over confidence, it would let model ignore the weak part in *softmax* matrix, so label smooth is introduced into the *softmax* function to lower the impact on over confidence that the *softmax* brings and let the model focus slightly on the weights of low probability distributions. The *softmax* with label smooth is shown in Eq.8, where *T* is a scalar temperature hyperparameter.

$$Softmax(x_i) = \frac{\exp\left(\frac{x_i}{T}\right)}{\sum_j^N \exp\left(\frac{x_j}{T}\right)} \tag{8}$$

Cross entropy is chosen as the loss function, as shown in Eq.9, where y represents the true label while $\hat{y}$ represents the predicted label and *n* is the number of samples.

$$L(\hat{y}, y) = -\sum_{i=1}^{n} y_i \log(\hat{y}_i) \tag{9}$$

Optimizer matters in backpropagation, we use *Adam* as the optimizer in model training, which is shown in Eq.10 to Eq.12, where β_1 and β_2 are constants that control the exponential decay, and m_t is the exponential moving average of the gradient, obtained by the first moment of the gradient. v_t is the squared gradient, obtained by the second moment of the gradient. Eq.7 presents the update process of the weight, where w_t refers to the weight at time t while w_{t-1} refers to the weight at time $t-1$. α is the learning rate of model training and σ is a non-zero constant.

$$w_t = w_{t-1} - \alpha * \frac{\widehat{m}_t}{\sqrt{\hat{v}_t + \sigma}} \tag{10}$$

$$\widehat{m}_t = \frac{m_t}{1-\beta_1^t} \tag{11}$$

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t} \tag{12}$$

C. Detection

The trade-off between time and accuracy has always been a problem in object detection, so the object detection network is divided into one-stage and two-stage. This paper chooses and fine-tunes 2 two-stage target detection networks including Faster R-CNN and DETR for RA radar target detection. Both Faster R-CNN and DETR consist of 2 parts, feature extractor and region proposal network (RPN). Feature extractor plays a role of extracting the semantic feature from the feature map while RPN takes the charge of finding the region of interest (ROI) and classify the target in ROI. The difference between them is at RPN, for Faster R-CNN uses CNN-based network while DETR uses vision-transformer-based network, which is worth comparing especially for the complex scene detection.

For the loss function, there are 2 parts, where L_{class} is the class loss and L_{bbox} refers to the bounding box loss.

For class loss, we still use cross entropy, which is shown in Eq.13, where c_i represents the ground truth label and $\emptyset$ represents no object(background), so $1_{c_i \neq \emptyset}$ is a Boolean function to identify whether the image contains the object or not. $-log\ p_i(c_i)$ is the standard cross entropy function, where p_i refers to the predicted label.

$$L_{\{class\}} = \sum_{\{i=1\}}^{N} -1_{c_i \neq \emptyset} log p_i(c_i) \tag{13}$$

The loss function of bounding box is shown in Eq.14. There are 2 items, the first is the loss of intersection over union (IoU), which is to calculate the loss value of the IoU between the ground-truth bounding box b_i and the predicted bounding box $\hat{b}_i$. The second item is the loss between the coordinates of the ground-truth bounding box and predicted bounding box. λ_{IoU} and λ_{L1} are 2 hyperparameters. Only the loss for coordinates cannot exactly measure the accuracy of the predicted bounding box, which is especially unfriendly to the small target, so IoU loss is added to balance and enhance the model confidence.

$$L_{bbox}(b_i, \hat{b}_i) = \sum_{i=1}^{N} \lambda_{IoU} L_{IoU}(b_i, \hat{b}_i) + \lambda_{L1} \left\| b_i - \hat{b}_i \right\|_1 \tag{14}$$

We still use Adam as the optimizer to train the object detection neural network, which is the same with that in classification.

V. EXPERIMENTS

The experiments can be divided into 2 subsections, classification and detection.

A. Classification

For classification, it is conducted on EA map and RA map respectively. There are 5 classes for both EA map and RA map. Each class has 1000 images, which are resized as 224 * 224(px). We split the dataset into a training set and a test set in a ratio of 9:1.

We use 4 models in total, including EfficientNet-b2, ResNet-50, ViT-small and TNT-small, 2 CNN-based and 2 vision-transformer-based neural network.

The hyperparameters in the training process can be seen in Table 2.

Table 2. Hyperparameters of Classification Network Training

Hyperparameter	Learning rate	Initial epochs	Batch size
Value	0.0001	30	16

Besides, we divide the data into 3 categories according to the classification difficulty. For the targets close to the radar and with high resolution, we divide them into simple cases; for the targets with medium distances from the radar, we divide them into medium cases; for the targets far from the radar and the target some of whose parts are outside of the radar angular field of view, we classify them as hard cases. Table 3 and Table 4 show the results of the classification on EA and RA map for all 4 models respectively.

It can be seen from Table 3 that the accuracy of two CNN-based models on easy cases are relatively closed to that of vision-transformer-based models. The gap between these 2 kinds of models is gradually apparent on the medium case. For the hard cases, the gap is obvious that TNT-small performs better than Efficient-b2 about 5%-6% while better than ResNet-50 about 5%-11%. From this perspective, multi-head self-attention mechanism in vision transformer could better recognize the targets which are more complex. In addition, compared with CNN-based network with its own inductive bias, vision-transformer-based network that abandons the inductive bias has stronger learning ability for irregular radar images with complex distribution rules and strong randomness, which can learn more potential features to improve the accuracy of the model.

Table 3. Classification Accuracy of 4 Models in EA Maps

Model	Easy Cases	Medium Cases	Hard Cases
ResNet-50	98.24%	91.19%	84.76%
Efficientnet-b2	96.10%	**94.19%**	86.77%
ViT-small	97.88%	93.71%	91.02%
TNT-small	**98.63%**	94.17%	**91.87%**

Table 4. Classification Accuracy of 4 Models in RA Maps

Model	Easy Cases	Medium Cases	Hard Cases
ResNet-50	83.69%	76.98%	74.99%
Efficientnet-b2	88.12%	84.74%	81.68%
ViT-small	**93.18%**	84.92%	82.97%
TNT-small	92.63%	**86.17%**	**83.33%**

To help intuitively understand the self-attention mechanism in the vision-transformer-based network, we develop a self-attention visualization module. Fig.4 shows 2 sets of self-attention maps of EA and RA maps respectively with ViT-small. The 7 self-attention maps in one set correspond to 7 self-attention heads, each self-attention head learn the different self-attention weights, but the target areas focused are approximately the same, which could further prove that after ViT flattens a two-dimensional image into a one-dimensional patch sequence, ViT could still learn the location information in the image from the image patch sequence without spatial information and highlight the important feature position. In addition, visualization of self-attention maps could also check whether the neural network is overfitting or not.

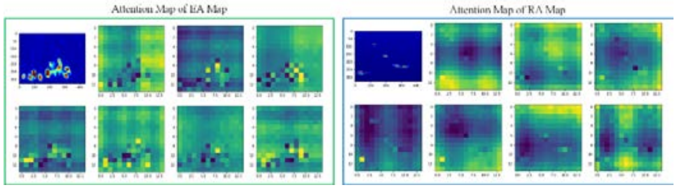


Fig. 4. Visualization of self-attention map with ViT-small

B. Detection

Detection is conducted only on RA map. In each RA map, there are one or more targets, which include traffic objects and accidents with different ranges, azimuths and poses randomly. There are 5614 RA maps and their annotation files in the dataset, which is organized as the format of *PASCAL VOC*. The class number of the targets is 5, which is the same with the experiment of classification. We split the dataset into a training set and a test set in a ratio of 9:1.

We respectively use Faster R-CNN(FPN) and DETR(PVT) as our object detection network. Stochastic gradient descent with momentum is chosen as the optimizer. The hyperparameters are shown in Table 5. In addition, early stopping is used in our experiment to avoid overfitting.

Table 5. Hyperparameters of Radar Object Detection Network in Training

Hyperparameter	Learning rate	Initial epochs	Batch size
Value	0.005	30	16

The examples of detection results are shown in Fig.5, where there are bounding box, predicted label and confidence probability score for each target.

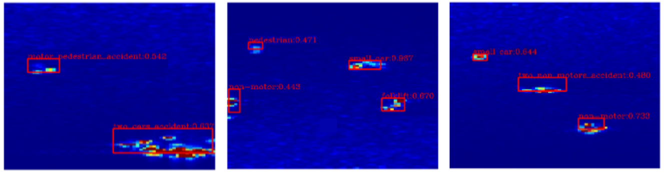


Fig. 5. Examples of detection results.

We select precision as the metrics, as Eq.15 shows:

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

where TP is true positive; FP is false positive; TN is true negative; FN is false negative. Precision measures the true detection rate while recall reflects the missed detection rate. In addition, we set 0.5 as the threshold for IoU. If IoU between predicted bounding box and ground-truth bounding box is greater than 0.5 and the classification result is true, the detection result is true positive. The detection result is shown in Table 6. We divide the detection result into 3 parts according to the different IoU between the predicted bounding box and ground-truth bounding box.

It is observed that DETR performs better than Faster R-CNN. The *mAP* of DETR is 1.84% higher than that of Faster R-CNN. For all the cases that IoU is above 0.50, the performances of Faster R-CNN and DETR are very close, but the gap becomes greater when the IoU is above 0.50.

Table 6. Detection Results of 4 Models on EA Map

Model	mAP	AP_{50}	AP_{75}	AP_{90}
Faster R-CNN	91.79%	96.82%	93.78%	84.76%
DETR	**93.63%**	**97.76%**	**96.94%**	**86.19%**

VI. CONCLUSIONS

The paper proposes an emulation platform to emulate the relatively complex traffic scenes like traffic accidents and develop an imaging FMCW radar module to generate the radar map of the target, which highly reduces the difficulty and cost of collecting data in the real world without danger. Secondly, the CNN-based neural network, as the conventional classification and detection solution, performs not as well as the vision-transformer-based neural network especially in the classification and detection of the complex scenes, for the vision-transformer-based neural network could learn richer context information of the complex targets with multi-head self-attention mechanism. However, the vision-transformer-based neural network dismisses the inductive bias, but the network can still learn the inductive bias, which is specific to the existing data. However, the real traffic accident scenes are far more complex than those included in our experiments. On the one hand, we need to increase the amount and type of traffic accident data in the future. On the other hand, as the era of 4D radar is coming, we will use 4D radar with high resolution and high detection accuracy to research event detection, where RCS and velocity should also be considered in the real scenes, especially for real-time detection. Last but not least, we will use the model pre-trained on the emulation dataset to train and test on the real data to improve the prediction performance.

ACKNOWLEDGEMENT

The authors acknowledge XJTLU-JITRI Academy of Industrial Technology for giving valuable support to this joint project. This work is also partially supported by the Xi'an Jiaotong-Liverpool University (XJTLU) AI University Research Centre, Jiangsu (Provincial) Data Science and Cognitive Computational Engineering Research Centre, and funding: XJTLU-REF-21-01-002.

REFERENCES

[1] Dong J, Chen S, Zong S, et al. Image transformer for explainable autonomous driving system[C]//2021 IEEE International Intelligent Transportation Systems Conference (ITSC). IEEE, 2021: 2732-2737.

[2] Okuda R, Kajiwara Y, Terashima K. A survey of technical trend of ADAS and autonomous driving[C]//Technical Papers of 2014 International Symposium on VLSI Design, Automation and Test. IEEE, 2014: 1-4.

[3] Chang S, Zhang Y, Zhang F, et al. Spatial Attention fusion for obstacle detection using mmwave radar and vision sensor[J]. Sensors, 2020, 20(4): 956.

[4] Major B, Fontijne D, Ansari A, et al. Vehicle detection with automotive radar using deep learning on range-azimuth-doppler tensors[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019: 0-0.

[5] Roldan I, del-Blanco C R, Duque de Quevedo Á, et al. DopplerNet: a convolutional neural network for recognising targets in real scenarios using a persistent range–Doppler radar[J]. IET Radar, Sonar & Navigation, 2020, 14(4): 593-600.

[6] Patel K, Rambach K, Visentin T, et al. Deep learning-based object classification on automotive radar spectra[C]//2019 IEEE Radar Conference (RadarConf). IEEE, 2019: 1-6.

[7] Fu R, Al-Absi M A, Kim K H, et al. Deep Learning-Based Drone Classification Using Radar Cross Section Signatures at mmWave Frequencies[J]. IEEE Access, 2021, 9: 161431-161444.

[8] Lim S, Lee S, Yoon J, et al. Phase-based target classification using neural network in automotive radar systems[C]//2019 IEEE Radar Conference (RadarConf). IEEE, 2019: 1-6.

[9] Zhao Z, Song Y, Cui F, et al. Point cloud features-based kernel SVM for human-vehicle classification in millimeter wave radar[J]. IEEE Access, 2020, 8: 26012-26021.

[10] Liu J, Zhang K, Sun Z, et al. Concealed object detection and recognition system based on millimeter wave fmcw radar[J]. Applied Sciences, 2021, 11(19): 8926.

[11] Cai X, Giallorenzo M, Sarabandi K. Machine Learning-Based Target Classification for MMW Radar in Autonomous Driving[J]. IEEE Transactions on Intelligent Vehicles, 2021, 6(4): 678-689.

[12] Bai J, Zheng L, Li S, et al. Radar transformer: An object classification network based on 4d mmw imaging radar[J]. Sensors, 2021, 21(11): 3854.

[13] Droitcour A D, Boric-Lubecke O, Lubecke V M, et al. Range correlation and I/Q performance benefits in single-chip silicon Doppler radars for noncontact cardiopulmonary monitoring[J]. IEEE Transactions on Microwave Theory and Techniques, 2004, 52(3): 838-848.

[14] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.

[15] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[16] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[17] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[18] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.

[19] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//European conference on computer vision. Springer, Cham, 2020: 213-229.

[20] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009: 248-255.

A Real-World Case Study of a Vehicle Routing Problem

Arnas Matusevičius[1], Rūta Juozaitienė[2], and Tomas Krilavičius[3]

***Abstract*— The goal of this study is to create a framework for route planning. The proposed approach considers the common features, i.e., picking up multiple freight according to the time-windows the pick-up and delivery locations have. However, a unique feature to the original Pickup-and-Delivery problem with time windows is introduced. Namely, freight can be redirected to depots for a fee, which lets drivers spend less time on the road and collect the redirected freights in one place. The genetic algorithm proves to be a viable approach as it produces reasonable results in a relatively short period of time.**

***Index Terms*— Vehicle Routing, Optimization, Genetic Algorithm, Nearest Neighbour, Pickup-and-Delivery.**

I. INTRODUCTION

Road freight transportation is rapidly expanding in a competitive environment, hence logistics companies with limited transportation capacities are forced to look for more efficient solutions that concern freight transportation. One of the key indicators used to determine a company's efficiency is the profit generated by the transportation services provided, which directly depends on the route taken. Usually, a logistics manager ensures high-quality, fast transportation of a customer's freight by striving to plan the best routes for truck drivers to take. In addition, the logistics manager must consider the available human resources, capacities of the vehicles, the order in which freight is taken and the expenses that will accumulate during each planned route. All of this is considered to ensure efficient management of transportation services while minimizing the company's expenses. Making sense of such a vast amount of information is not a straightforward task, hence, it is normally formulated as an optimisation problem.

Optimization of routes regarding various constraints is known in scientific literature as the Vehicle Routing Problem (VRP). To create a framework for road freight transport, optimization methods applied to the Pickup-and-Delivery (PDP) are analyzed as the problem analzed in them matches our case study best. In this case, it is assumed that each load has a predefined delivery point (one-to-one) and that the goods can only be picked up and delivered at certain hours (time windows).

Also, a possibility to redirect freight to depots was added. Depots use trucks and/or buses to collect freight to its destination.

If a freight is transported to a depot it is later picked up by a truck from the location of the depot. That way, the truck does not need to travel to the pick-up location. In addition, several freight from different locations may arrive at a depot, which can later be simultaneously picked-up by a truck. This usually significantly reduces the distance and time needed visit all of the freight pick-up locations. However, the depot services come with a charge, hence it is important to consider the situation before making a decision. Each load has an assigned a delivery rate to the depot, based on the weight of the load and the distance to the depot. When analyzing the possibility of transporting the freight through the depot, the algorithm selects the depot with the lowest service fee for the freight in question. Such problems are often unsuitable to linear programming methods due to extremely long computation times, therefore a metaheuristic approach is used.

The rest of the paper is organized as follows. Related work in this area is presented in section II. Section III introduces the datasets used in the current study. Section IV presents the proposed approach to this problem. Experimental results are provided in section V. Finally, concluding remarks and plans for the future are discussed in Section VI.

II. RELATED WORK

Vehicle routing problems originated from the generalization of the Traveling Salesman Problem (TSP). The simulated annealing [1] and Tabu search [2], [3] algorithms were tested to solve it. However, experiments revealed that these methods require large computational time resources. Also, a genetic algorithm method has been proposed to solve this type of problem. For example, a genetic algorithm has been used to create school bus routes [4]. This study revealed that fairly good results were obtained in a relatively short time.

The creation of transportation routes with time windows (VRPTW) where customers can be served only for a specified time interval is analyzed in [5]. The author proposed using a genetic algorithm to determine how many cars are needed and a Tabu search algorithm – to reduce the total distance travelled by cars. The author notes that using both algorithms is more suitable for this (VRPTW) problem, rather than using a single of the aforementioned algorithms.

In [6] a Pickup-and-Delivery with time windows (PDPTW) problem is analyzed. The study has shown that dynamic programming is not suitable for solving this problem due to the long computation time, whereas the results obtained by a

Authors 1,2 are with the Department of Mathematics and Statistics and author 3 is with the Department of Informatics, Vytautas Magnus University, Kaunas, Lithuania. (email: arnas.matusevicius@vdu.lt, ruta.juozaitiene@vdu.lt, tomas.krilavicius@vdu.lt).

genetic algorithm were able to provide (sub)optimal solutions for problems bigger by up to 25% of the original problem.

In [7] study, a PDPTW problem where not all goods are required to be transported is analyzed using a hybrid genetic algorithm. The genetic algorithm would take a few minutes to produce a good and stable result, whereas linear programming methods took more than two hours to reach these results. Also, the same problem was studied in [8]. Here, three metaheuristic methods were suggested, namely the Tabu search, the genetic algorithm, and the scatter search. Although all methods provided good results, the quality of the Tabu search results as well as the speed of convergence was notably better.

Taking everything into consideration, we can state that the problem solved in this case study belongs to the group of complex combinatorial problems. Linear programming methods take too long to solve such problems, so although they are suitable in theory, they are not implemented for solving real-world problems. Various scientists solve this problem using metaheuristic methods, of which the most popular was found to be the genetic algorithm approach.

III. RESEARCH DATA

The research was conducted with real historical data. It consists of information about freight transported from January 2015 to December 2021. The number of completed orders is 167 697 and the number of freights is 177 525. All information about these orders is stored in data tables:

1) CargoOrder -- information about each order.
2) CargoLoad -- information about each load of an order.
3) CargoOrderStage -- information about the stages of the trip.
4) ExpediteTrip -- information about the trip.

Based on this data, a data set was created and used for the experimental study of the methods. The list of dataset attributes includes information about:

1) Order ID.
2) Freight ID.
3) Revenue earned per freight.
4) Sender's address.
5) Sender's city.
6) Sender's country.
7) Sender's postal code.
8) Receiver's address.
9) Receiver's city.
10) Receiver's country.
11) Receiver's postal code.
12) Date when the order was created.
13) Sender's coordinates.
14) Receiver's coordinates.
15) Freight's loading metres (LDM).
16) Freight's weight.
17) Freight's volume.
18) Information about whether the freight was redirected to a depot in the sender's country.
19) Information about whether the freight was redirected to a depot in the receiver's country.
20) Information about whether the freight had an intermediate stop in another country.
21) Coordinates of intermediate stops.

IV. PROPOSED APPROACH

Case study of a PDP problem

The object of this research is to create a route planning algorithm in such a way that the profits would be maximized and that it would have these features:

- **able to process large amounts of data.**
- **ensure that the capacity of a truck will not be exceeded.** Neither one of the three given capacity dimensions can be exceeded when planning which freight to pick up.
- **assess which goods are not profitable and remove them from planning.**} Freight that is worth less than what it would cost to transport it, should be removed from planning.
- **able to redirect freight to depots when it is more convenient.** Any freight can be redirected to a depot which will collect the assigned freight for a fee. Usually, a freight gets redirected to a depot with the lowest price. Redirecting freight to a depot is usually done either to minimize the distance driven by a driver or to minimize the collection time of the freight, as each freight has a delivery deadline that when exceeded causes additional expenses.
- **consider the working hours of each location.** When planning a route, the work hours of pick-up locations need to be considered as well as when each freight is ready to be picked up.
- **estimate the cost of delays and downtime (caused by reaching a pick-up location too early, as the goods are considered to be not ready for transport or a delivery location outside of its work hours).**
- **able to use different time zone data.** Since the algorithm creates international routes, it should be capable of assessing the deadlines and work hours of establishments in different time zones.
- **ensures that drivers' hours of service are not exceeded.** Mandatory breaks from driving must be considered when assessing the time, it takes to get from one location to another. The Hours of Service (HoS) used in this research state that: a 45-minute break must be done after 4.5 hours of consecutive driving, an 11-hour break after driving 9 hours per day and a 45-hour break after driving 56 hours per week.

Model formulation

A total of n transportation requests are represented as a directed graph $G = (V, A)$; where V is divided into nodes $P = \{1, \dots, n\}$ for pickup, $D = \{n + 1, \dots, 2n\}$ nodes for delivery and $Term = \{Term^1 \cup \ldots \cup Term^T\}$ for depot nodes, where T is the number of depots. Each freight i needs to be transported from node $i \in P$ to a delivery node $n+i \in D$. Freight is measured in three different dimensional characteristics, namely weight (w), volume (v) and loading meters (l) which we will denote as q_i^w, q_i^v, q_i^l accordingly. Let $q_{n+1}^w = -q_i^w, q_{n+1}^v = -q_i^v, q_{n+1}^l = -q_i^l$. Also, each freight brings revenue e_i that is known beforehand. The depots charge differently according to the weight and distance that needs to be driven, hence each depot has different transportation prices for the freight $Term^j = \{c_1^j, \ldots, c_n^j\}$, where c_i^j is the price of i-th freight in the $j$-th depot. Therefore, we introduce a binary decision variable x_{ij}^k that is equal to 1 if freight i of the vehicle k is redirected to a depot $j$ and zero otherwise. After redirecting the freight into a depot, the pickup location of the freight i will change into the location of the depot $j$ and the transportation price of the depot $c_i = \min\{c_i^1, \ldots, c_i^T\}$ is added to the total expenses. Furthermore, each node $i \in V$ must be visited within a time window $[a_i, b_i]$. A visit requires a certain time s_i to process. The time a kth truck starts servicing at node i is denoted by T_{ik} and $Q_{ik}^w, Q_{ik}^v, Q_{ik}^l$ denote the dimensions of the truck after servicing the i-th node. Another binary decision variable x_{ijk} is equal to 1 if the kth truck drives from node i to node $j$. Expenses related to the cost of transportation through each arc $(i,j) \in A$ are denoted as c_{ij} and have a duration of t_{ij}.

Moreover, it was noticed that sometimes the route would go straight through water or the borders of neighbouring countries, hence the addition of a borderland zone was developed. To reach the starting country's border point, a vehicle in the borderland zone would have to first visit the special location from which the route to the border is undisturbed. The borderline zone is coloured red in Fig. 4 and the special location is visualized by a black star icon.

The mathematical model can be formulated as follows:

$$\max \sum_{k\in K}\sum_{i\in V}\sum_{j\in V}\Bigg(x_{ijk}\,\mathbb{1}_P(j)p_j\,e_j + x_{ijk}\,\mathbb{1}_P(Term)\sum_{l\in P} x_l^j\,(e_l - c_l) - c_j^D\,\Delta_j - c_{ij}x_{ijk}\Bigg) \qquad (2.1)$$

$$\sum_{k\in K}\sum_{j\in V} x\_ijk = 1 \qquad \forall\, i \in P \cup Term, \qquad (2.2)$$

$$\sum_{j\in V} x_{ijk} - \sum_{j\in V} x_{n+i,jk} = 0 \qquad \forall\, i \in P, k \in K, \qquad (2.3)$$

$$c_i = \min\{c_i^1, \ldots, c_i^T\}, \qquad \forall\, i \in V \qquad (2.4)$$

$$T_{jk} \geq (T_{ik} + s_i + t_{ij})x_{ijk} \qquad \forall\, i \in V, j \in V, k \in K \qquad (2.5)$$

$$Q_{jk}^w \geq (Q_{ik}^w + q_j^w)x_{ijk} \qquad \forall\, i \in V, j \in V, k \in K, \qquad (2.6)$$

$$Q_{jk}^v \geq (Q_{ik}^v + q_j^v)x_{ijk} \qquad \forall\, i \in V, j \in V, k \in K, \qquad (2.7)$$

$$Q_{jk}^l \geq (Q_{ik}^l + q_j^l)x_{ijk} \qquad \forall\, i \in V, j \in V, k \in K, \qquad (2.8)$$

$$T_{n+i,k} - T_{ik} - s_i - t_{i,n+1} \geq 0 \qquad \forall\, i \in P, \qquad (2.9)$$

$$a_i \geq T_i \qquad \forall\, i \in P, Term \qquad (2.10)$$

$$a_i \leq T_{ik} \leq b_i \qquad \forall\, i \in V, k \in K, \qquad (2.11)$$

$$\max\{0, q_i^w\} \leq Q_{ik}^w \leq \min\{Q_k^w, Q_k^w + q_i^w\} \qquad \forall\, i \in V, k \in K \qquad (2.12)$$

$$\max\{0, q_i^v\} \leq Q_{ik}^v \leq \min\{Q_k^v, Q_k^v + q_i^v\} \qquad \forall\, i \in V, k \in K \qquad (2.13)$$

$$\max\{0, q_i^l\} \leq Q_{ik}^l \leq \min\{Q_k^l, Q_k^l + q_i^l\} \qquad \forall\, i \in V, k \in K \qquad (2.14)$$

$$x_{ijk}, x_i^t, p_i \in \{0,1\} \qquad \forall\, i \in V, j \in V, k \in K, \qquad (2.15)$$

$$\Delta_i \geq 0 \qquad \forall\, i \in V. \qquad (2.16)$$

The smallest depot price is ensured in (2.4) constraint. (2.5) equation states that the departure time at node $j$ must be later than the departure time at node i plus travel and processing time if route (i,j) is traversed. The consistency of load variables is ensured (2.6-2.8) constraints. Equation (2.9) introduces precedence constraints. Then, the (2.10) constraint ensures that the freight is picked up only when it is ready to be picked up. Furthermore, (2.11) constraint impose time-window and (2.12-2.14) capacity constraints, respectively.

Genetic algorithm

Genetic algorithms (GA) are widely used evolutionary computing methods for solving complex optimization problems that do not have the usual mathematical properties such as continuity, differentiability and convexity. These properties are often not satisfied in practical problems, so genetic algorithms are used to find their solutions, for which such requirements are not necessary [3].

Genetic algorithms are based on the principle "only the fittest survive". These algorithms work with entities, sometimes called chromosomes, each of which represents a possible solution to a problem. Typically, genetic algorithms start off by creating an initial set of solutions called a population. In each subsequent iteration, GA creates a new potential offspring solution from the parent solutions selected in the previous generation using crossover and mutation operations. This process of generating and evaluating solutions continues until predefined criteria are met [4].

k-Nearest Neighbours

Constructing the initial population is crucial in using genetic algorithms as it directly affects the time it takes to converge. An initial population closer to the real solution leads to a faster (sub)optimal solution finding. For this study, the nearest neighbour (k-NN) algorithm is used for calculating the initial population.

The k-nearest neighbor method is based on learning by analogy [9]. The training samples are described by n-dimensional numeric attributes. Each sample represents a point in an n-dimensional space. In this way, all of the training samples are stored in an n-dimensional pattern space. When given an unknown sample, the ***k*-nearest neighbor method** searches the pattern space for the k training samples that are closest to the unknown sample. These k training samples are the k "nearest neighbors" of the unknown sample. "Closeness" is defined in terms of Euclidean distance, where the Euclidean distance between two points, $X = (x_1;\ x_2;\ \dots;\ x_n)$ and $Y = (y_1;\ y_2;\ \dots;\ y_n)$ is:

$$d(X,Y) = \sqrt{\sum_{i=1}^{n}(x\_i - y\_i)\text{^}2}\,.$$

The unknown sample is assigned the most common class among its k nearest neighbors. When $k = 1$, the unknown sample is assigned the class of the training sample that is closest to it in pattern space.

V. RESULTS

After finding the initial population, the genetic algorithm, depicted in Fig. 1, is initialized.

The effectiveness of the genetic algorithm is affected by the selected model parameters. To increase the genetic algorithm's efficiency, we performed a simulation study comparing different operators: for the selection component we compared lhe linear-rank (lr), the nonlinear-rank (nlr), the proportional/roulette wheel (rw) and the unbiased/tournament (tour). As for the crossover component, the cycle (cx), the partially matched (pmx), the order (ox) and the position-based (pbx) crossover operators were compared. What concerns the mutation component, the simple inversion (sim), insertion (ism), exchange/swap (sw), displacement (dm) and scramble (scr) mutation operators were compared. In Fig. 2 the summarized results of 50 replications are displayed. Values in

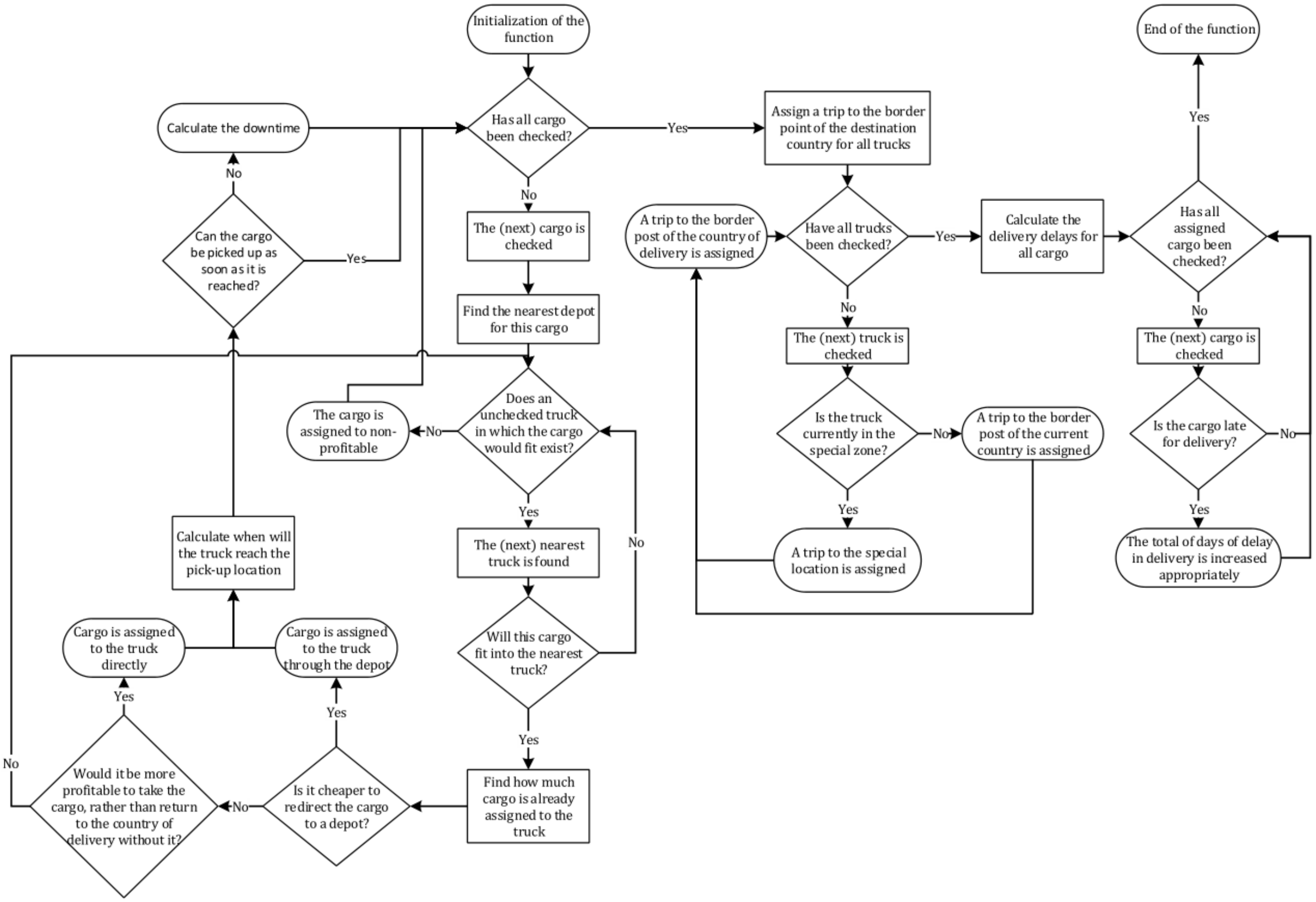

Initialization of the function
Has all cargo been checked?
Yes
No
Assign a trip to the border point of the destination country for all trucks
End of the function
Calculate the downtime
No
Can the cargo be picked up as soon as it is reached?
Yes
The (next) cargo is checked
Find the nearest depot for this cargo
A trip to the border post of the country of delivery is assigned
Have all trucks been checked?
Yes
Calculate the delivery delays for all cargo
Has all assigned cargo been checked?
Yes
No
The (next) truck is checked
No
The (next) cargo is checked
The cargo is assigned to non-profitable
No
Does an unchecked truck in which the cargo would fit exist?
Is the truck currently in the special zone?
No
A trip to the border post of the current country is assigned
Is the cargo late for delivery?
No
Yes
Yes
Yes
Calculate when will the truck reach the pick-up location
The (next) nearest truck is found
No
A trip to the special location is assigned
The total of days of delay in delivery is increased appropriately
Cargo is assigned to the truck directly
Cargo is assigned to the truck through the depot
Will this cargo fit into the nearest truck?
Yes
Yes
Yes
No
Would it be more profitable to take the cargo, rather than return to the country of delivery without it?
No
Is it cheaper to redirect the cargo to a depot?
Find how much cargo is already assigned to the truck

Fig. 1. Route planning and constraint checking scheme

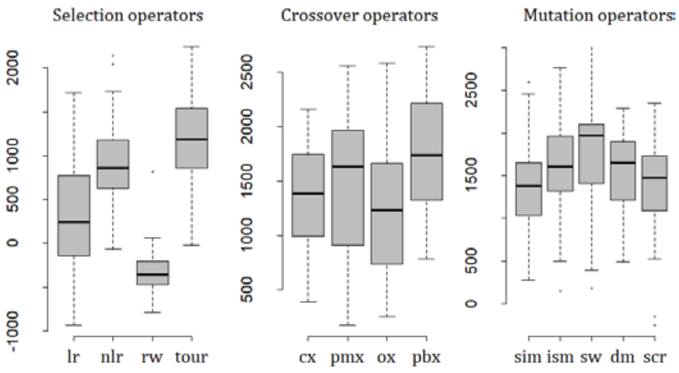


Fig. 2. Comparison of different genetic algorithm operators

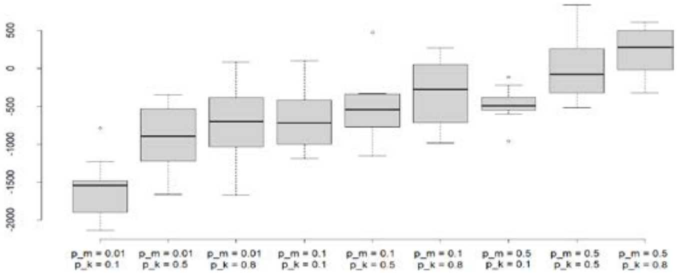


Fig. 3. Comparison of values generated with different crossover and mutation probabilities

the Y-axis denote the total profit of a tour; hence a higher value means a more profitable route created by the algorithm.

Furthermore, p_m denotes the probability value of the mutation operation and p_k the probability value of the crossover operation. The results reveal that the best results are achieved with the tournament selection, position-based crossover, and swap mutation operators. Also, experiments were performed to determine favourable probabilities between crossover and mutations pairs in the parental chromosome. The results presented in Fig. 3 show, that using the probability of crossover p_k equal to 0.8 and the probability of mutation p_m equal to 0.5, yields the highest profits.

The simulations were carried out using a real data set, which describes freight situated in the territory of Germany in the period 15/11/2021 -- 19/11/2021. This dataset consisted of 6 trucks and 38 freight. In the beginning all trucks are considered to be identical and empty, i.e., all trucks have a carrying capacity of 24 000 kg, 120 m^3 and 13.6 m (LDM). Also, since the initial positions of the trucks are not stored in the historical data, their initial locations were chosen to match the locations of the biggest freight LDM-wise in the data. The main focus of this analysis was on the collection of freight. Therefore, the freight delivery route was not planned. However, a delivery of all freights to a fixed final point (the Lithuania Poland Border Crossing was set as the final point) was implemented.

In Fig. 4, a visualization of the results is presented. All freight is depicted as either coloured dots (colours depend on which truck it is assigned to) or diamonds. Diamond shaped locations are freight redirected to the depot (a square of the same color is the depot to which it is assigned to). Different colored lines represent routes of individual trucks. In this case, a borderland zone is constructed around the bottom of Germany as to ensure that the truck does not enter The Czech Republic. The black star near the west-most of The Czech Republic is the special

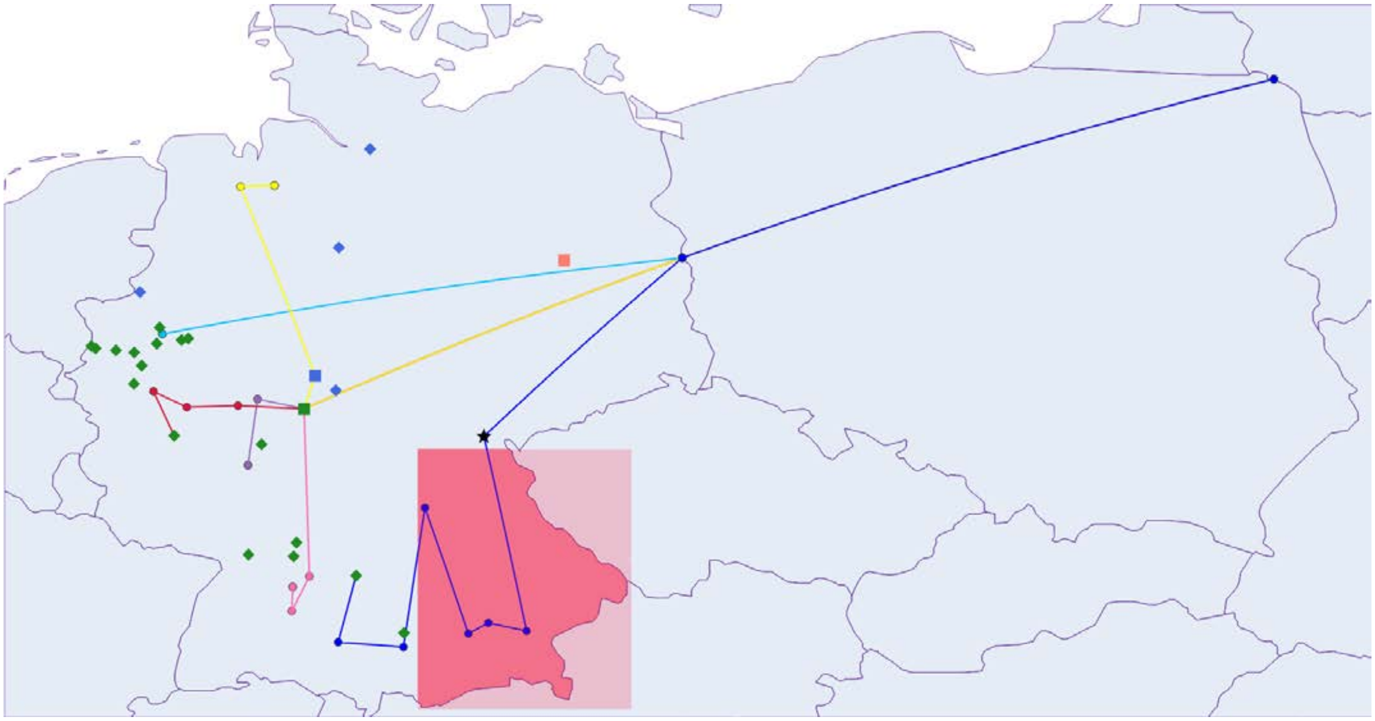


Fig. 4. Visualization of the results

location that trucks would need to pass before continuing their drive to the border.

The created algorithm obtained the (sub)optimal solution after performing 70 iterations and the calculations lasted 7 minutes and 20 seconds. Time constraints prolong the calculations as the working hours of all the trucks, break times, the time during which the distances will be covered, the working hours of the locations and the deadlines for picking up/delivering the freight has to be tracked. For example, without time constraints, the algorithm obtained an (sub)optimal solution after performing 320 iterations (with a duration of $\sim$ 5 minutes and 50 seconds), i.e., the computation time of a single iteration increased as much as ~ 7 times.

Moreover, the route created using the 1-nearest neighbour method had a total of 11 796.76 km distance driven and a profit of -3 614.57 €. Whereas the results calculated using the GA approach concluded with a total driven distance of 11 149.49 km and profits of -369.58 €. The GA approach proves to be more successful as it looks into more than the distance between a freight and a truck.

VI. CONCLUSIONS

The addition of freight redirection to depots greatly decreases the time it takes to collect all freights assigned to a truck. Therefore, less freight delivery deadlines are exceeded, and more profits are kept. Also, simulations with different GA operators and their probabilities revealed, that the best results are achieved with the tournament selection, position-based crossover (probability $p\_k = 0.8$), and swap mutation (probability $p\_m = 0.5$) operators.

Furthermore, experiments with real world data conclude that the genetic algorithm approach is suitable and should be used to create (sub)optimal transportation routes. However, the addition of time constraints significantly increases the calculation time of each iteration. Nonetheless, the proposed approach produces logical truck routes in a short period of time.

To improve the algorithm's performance, we plan a two-step approach, which reduces the complexity of the problem, hence diminishing the computation time. The first step is described in this paper, and the second step, that is planned for future, would consist of delivering all of the gathered freight at the border point to their delivery destinations.

VII. ACKNOWLEDGEMENTS

We thank Egidijus Grigas and Donatas Čelkys from UAB TerraIT, and Egidijus Lučka from UAB AD Rem for cooperation and useful insights.

REFERENCES

[1] G. B. Alvarenga, G. R. Mateus, and G. De Tomi, A genetic and set partitioning two-phase approach for the vehicle routing problem with time windows, Computers & Operations Research, vol. 34, no. 6, pp. 1561–1584, 2007.

[2] L. Za, M. Suroso, I. Astuti, D. Khairina, and S. Maharani, Application of haversine formula in education game “landmark nusantara”, 01 2021.

[3] J. Tao, R. Zhang, and Y. Zhu, Dna computing based genetic algorithm, applications in industrial process modeling and control, 01 2020.

[4] G. Vaira, Genetinis algoritmas transporto maršrutų sudarymo uždaviniams spręsti, doctor's disertation, Vilnius university, 2014.

[5] J. Han, J. Pei, and M. Kamber, Data mining: concepts and techniques. Elsevier, 2011.

[6] J. Brownlee, Clever algorithms: nature-inspired programming recipes. Jason Brownlee, 2011.

[7] E. K. Burke, E. K. Burke, G. Kendall, and G. Kendall, Search methodologies: introductory tutorials in optimization and decision support techniques. Springer, 2014.

[8] S. Luke, Essentials of Metaheuristics, 2nd ed. Lulu, 2013. [Online] Available: http://cs.gmu.edu/∼sean/book/metaheuristics/

[9] J. Han, J. Pei, and M. Kamber, Data mining: concepts and techniques., Elsevier, 2011.

User fears and challenges in the adoption of network automation

Deepika BR[ab]*, Woon Kian Chong[a] and Gert Grammel[b]

Abstract—**New and upcoming automation technologies promise communication service providers (CSP) the capacity to accelerate labor-intensive manual processes of their network operating centers (NOC), providing opportunities for network operators to achieve superior performance in the network and giving time for innovative ideas in their own NOCs. There are challenges for both the NOC leaders as well as the operators as they devise strategies to implement wide-spread automation, and there may be many different aspects to the slowness in the adoption of network automation. In this research, data from 43 different organizations has been gathered to understand the fears and challenges faced by the network operators in adopting automation. The analysis recognizes challenges encountered while implementing well-known technologies and concepts that are rapidly gaining popularity. The analysis also reveals that network operators have a variety of concerns, ranging from putting their trust in machines to difficulties in demonstrating a concrete return of investment (ROI) while adopting the technology.**

I. INTRODUCTION

CSPs are under a lot of pressure to modernize their networks due to the emergence of edge clouds, the growing acceptance of cloud-based services, and the rollout of 5G technologies and services. There's a great need for enterprises and CSPs to be nimble, flexible, resilient, and highly available to satisfy the expectations of the data-hungry services.

In addition to the rising need for the service providers to improve the quality of services, bandwidth, latency, and performance, these organizations are also under pressure to meet their bottom lines and are vying to meet the customer expectations while continuing to be profitable.

Network automation makes it easier to configure, run, and manage the lifespan of network services, enabling autonomous operations and thereby makes it vital for all businesses with medium size to big networks to make it a top priority.

Today, organizations are well aware that network automation is a necessity and are aware of the automation's potential to save operating expenses and boost profitability. However, in contrast to the relevance and popularity automation is gaining, there seems to be a lag in its implementation. Understanding the mindset of the consumers is crucial for firms who are looking to accelerate the adoption of the automation.

Following these lines, this paper aims to recognize the state of network automation in the industry, identify the fears of users while implementing automation that might be causing the slowness in the adoption of network automation, and devise methods to improve the confidence or trust of network operators in automation.

In this study, a non-normative approach is considered to examine how a community of network operators view automation and the anxieties about implementing it in their networks.

II. LITERATURE REVIEW

As detailed in the article by [1], since the days of Advanced Research Projects Agency Network (ARPANET), the Internet has undergone a massive development, and today the number of applications required to satisfy the demands of different types of deployments as well as user services are throwing considerable number of challenges to the network operators [1]

It is challenging to anticipate that a single, non-intelligent network will be dependable, adaptable, extensible, secure, and economical. Network operators are charged with setting up numerous 'top-notch' policies that will be used to manage the network [2]. Operators face the challenge of individually implementing all the necessary low-level and proprietary commands on each node separately in order to configure these multiple top-notch policies [3]. When it comes to heterogenous networks, this can get even more challenging. This calls for the network to possess intelligent capabilities and the ability to carry out a number of balancing tasks automatically.

Considering all the factors mentioned above, the current networks are facing unprecedented challenges in terms of network management and operational efficiency due to the lack of automation, which cannot be disregarded.

Network automation particularly is not new, but rather has been an evolution. Research from as early as 2003 in implementing an intelligent knowledge plane for the internet has been conducted to establish the importance of the network automation and intelligence. Intent-based networking (IBN), Autonomous Driving Networking (ADN), Intent-driven networking (IDN), Zero-touch Network (ZTN), self-driving network, and various other automation frameworks been put forth by organizations such as Cisco, Juniper, Huawei, ETSI, and others.

The following sections explore the benefits and challenges witnessed in some of the key stages of network automation evolution:

[a] SP Jain School of Global Management
[b] Juniper Networks, Inc.

SOFTWARE DEFINED NETWORKING (SDN): In most traditional networks, various hardware elements are distributed with the decision-making capacity or network intelligence. Because each of the various network nodes must be reconfigured, adding a new network device or service is a laborious task. This has made legacy or traditional networks very hard to automate [4]. Additionally, the usage of IP address to identify the devices does not go very well in large, virtualized networks.

SDN promises to help traditional networks by predominantly recommending the separation of the control plane and the data plane, thereby enabling the underlying devices to utilize the control plane capabilities to make decisions and the data plane capabilities to mainly forward the traffic.

Fig. 1 illustrates the SDN controller managing the control plane aspects and separating it out from the data plane layer.

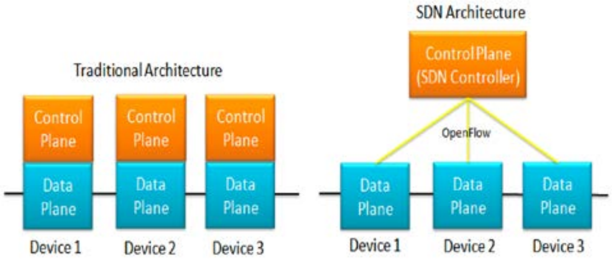


Fig. 1. Traditional Architecture vs. SDN Architecture

From Qin, Z., Denker, G., Giannelli, C., Bellavista, P., & Venkatasubramanian, N. (2014). A Software Defined Networking architecture for the Internet-of-Things. 2014 IEEE Network Operations and Management Symposium (NOMS). https://doi.org/10.1109/noms.2014.6838365

Although SDN brings a number of benefits such as ease of configuration and troubleshooting, performance improvement, cost savings, improved programmability, SDN also comes with several challenges. Some of the main challenges of SDN implementation include:

Reliability and Scalability: In the eventuality of a device failure in a manually controlled network, network operators reroute the traffic to alternate paths and alternate devices manually. SDN controller being centralized, takes up the role of the complete decision-making, thereby posing the challenge of letting down the entire network in the eventuality of the controller failure [5]. The controller must therefore support redundancy, via clustering of two more controllers in active-active, or active-standby mode [6].

Additionally, as the number of nodes in the network increase, the decisions that must be taken in order for the traffic to switch also has to be executed rapidly. The controller being a single point of control must possess robust performance, failing which, the number of requests that reach the controller can queue up, resulting in the slowness of the entire system.

Interoperability and compatibility: One of the major challenges that network operators are facing moving to an SDN approach is to get all the devices in the network to work with each other, and then with the controller. This is especially true in networks that have many legacy equipment. It is also important that when multiple controllers are working in order to control a set of devices in the same domain or multiple domains, these controllers are compatible with each other [6]. Although several standard bodies such as Internet Engineering Task Force (IETF), Open Networking Foundation (ONF), and others are constantly working towards providing standardized methods of network implementations, it will take a while before all vendors start implementing and comply to standards in their devices.

Security: Different SDN controllers can come with different security feature implementation. The security aspects that are already present in the current or the legacy networks, with respect to the security protocols, devices, and features implemented may not be supported by the SDN controllers. The SDN controllers being a single point of intelligence also increases its vulnerability towards hackers' attacks [7].

There are also several other challenges such as the programming interface between the controller and network devices, placement of the controller, SDN security applications and standardization of SDN that are causing network operators to be wary of implementing these solutions.

CLOSED CONTROL LOOPS: As much as the network is being made simpler by introducing virtualization at every level, it is also bringing more challenges to manage the network. It is therefore important to understand the significance of virtualization and the changes it brought along in the domain of automation.

The interest that network operators have shown in virtualization is one of the key factors that has influenced the architecture of 5G networks [8]. Virtualization facilitates switching to commercially available off-the-shelf (COTS) servers, promising to bring lower capex and demand-driven resource allocation, as well as the ability to maintain network slices. The flipside of these benefits is the complication that is introduced in managing and maintaining the operations. Automation therefore becomes an imperative in managing beyond-5G networks. In this case, automation is not just expected to fulfil the basic capabilities but is also expected to facilitate Artificial Intelligence (AI) and Closed control loops (CCLs) for functioning across different technology boundaries in a multi-vendor context.

[8] define the concept of CCLs in detail. With CCLs, once the goal of the network has been defined, there's an expectation to run all the stages of the operations starting from planning through analysis without any operator participation. **Fig 2** illustrates the expectations from complex and interconnected CCLs. Although CCLs carry a number of benefits and provides a vision of self-driven, self-correcting, and self-sufficient networks, they do carry challenges too.

Following are some of the key challenges identified in implementing CCLs:

Managing and controlling CCLs: When managing end-to-end (E2E) services through CCLs, it is key that the entities that are outside of this CCL are aware of how a particular set of services are being managed and are integrated accordingly.

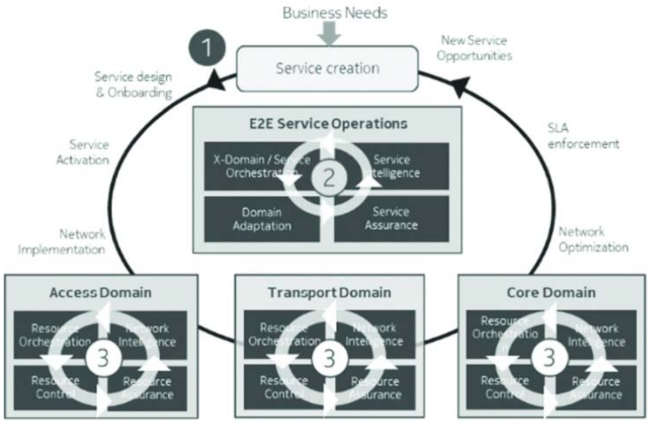


Fig 2 Complex and interconnected CCLs
From Vaishnavi, I., & Ciavaglia, L. (2020, November 1). Challenges Towards Automation of Live Telco Network Management: Closed Control Loops. IEEE Xplore

Operator's trust in CCL: Operators hesitate to enable the CCLs in a production system for the fear of the CCL taking the complete control of the network, thereby not allowing the operators a chance to look into the network in case of eventualities.

Standards for operational policies and inter-CCL operations: CCLs leverage the operational policies to obtain the instructions needed to operate and take actions such as remediations, workflows, delegation and so on as the changes are detected. When the CCLs are unable to take actions beyond a boundary, they escalate the outcome with the relevant information such as logs, and actions taken to the next level. There are also challenges associated with how to manage peer and hierarchical CCLs and in understanding how the CCLs influence each other. Standards such as ETSI GS ZSM0009-2 and ETSI GS ZSM0009-1 are working to improve these concepts.

AUTOMATION FOR 5G NETWORKS: A highly adaptable network design that supports E2E network slices spanning from the Radio Access Network (RAN) to the mobile core is necessary to support the complex requirements of 5G and future mobile networks [9]. The new architecture calls for decoupling of the traditional RAN architecture to Remote Radio Unit (RRU), Central Unit (CU), and Distributed Unit (DU).

Similarly, Next Generation (NG) core separating Evolved Packet Core (EPC) functions into much more intricate network functions helps in deploying both the vRAN and core NFs on COTS servers using VMs.

The virtual network is expanded by an E2E network slice into several technological and may be administrative network pieces. To achieve and realize a complete zero-touch orchestration and administration, closed-loop automation (CLA) is much needed.

Fig 3 depicts a CLA where every component can leverage AI/ML for intelligent monitoring of large amounts of data collected from physical and virtual devices and to provide an optimal monitoring frequency especially for network slices.

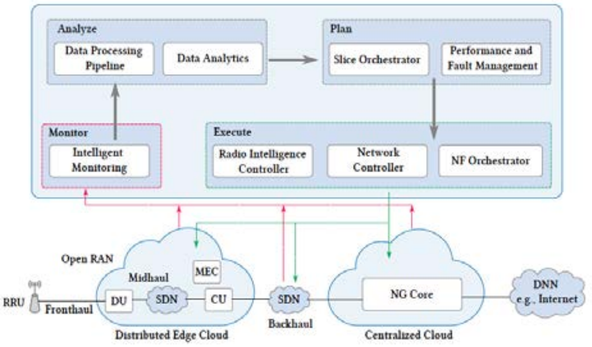


Fig 3 Closed loop management and orchestration
From Boutaba, R., Shahriar, N., Salahuddin, M. A., Chowdhury, S. R., Saha, N., & James, A. (2021). AI-driven Closed-loop Automation in 5G and beyond Mobile Networks. Proceedings of the 4th FlexNets Workshop on Flexible Networks Artificial Intelligence Supported Network Flexibility and Agility

As always, despite the benefits the automation framework brings here, there are several challenges associated with 5G automation:

Tracking slicing requirements: Every network slice comes with a set of unique requirements including latency, throughput, Bit Error Rate (BER) and so on. Mapping these requirements to the underlying infrastructure, such as RAN and core network function requirements, and isolating slices from one another and the resources or bandwidth consumed is very challenging [9].

Since the relationship between resource requirements and slice QoS requirements is non-linear and a number of factors, including network health, the degree of isolation between slices, and traffic volume, affect how end users perceive QoS, it is challenging to represent resource needs as a function of slice QoS requirements. Also providing a common orchestration platform for Disaggregated RAN by bringing all the different vendors involved in building the RAN ecosystem is non-trivial.

Fault and performance management: The ever-so-crucial fault and performance management comes with its own set of challenges whether it is proactive or reactive. Diagnosing the root cause of the problem, isolating alarms and establishing meaningful dependencies, and eliminating the recurrence requires the application of mitigation workflows. The most crucial challenge in all of this is the selection of the workflow that must be run when fault and performance degradation is detected. The idea of a bespoke *if-this-then-else-another* method is absolutely unworkable given the level of complexity that 5G networks entail.

There are many more forms and frameworks of automation, however, the idea of this paper is to explore the top few automation stages, understand the benefits, and also recognize the challenges that's causing slowness in the adoption.

]

III. RESEARCH METHODOLOGY

To understand the impact of the network automation growth versus the implementation in service provider organizations, a survey of about 20 questions was prepared and circulated to collect information from the network operators in the period of August to September 2022. The questions focused on understanding the network operators' approach towards network automation (benefits and drawbacks), possibilities of automation in their current job role, and their fears or apprehensions on automation. About 38 network operators provided responses to the questionnaire. Participation in this survey was voluntary and completely anonymous (which gave them the ability to express their opinions in an unbiased manner). Additionally, one-to-one interviews with the principal network operators of service provider organizations were conducted to obtain a more narrative understanding on their viewpoints and challenges of automation.

Results from the survey method

Manual vs. Automated operations

Most people rated running network automation manually was not a major task for them, and that they were either neutral or quite comfortable running operations manually. **Fig 4** shows the responses on comparison between manual versus automated operations.

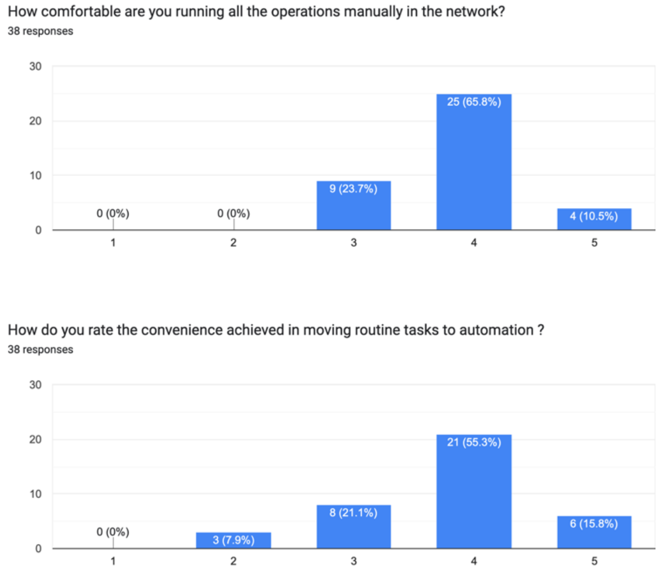


Fig 4 Comparison between manual vs. automated operations

However, when we take a look at how they perceived moving operations to automation, more than 70% rated it high indicating moving to automation can actually make things easier for them.

This shows that while they may be very used to managing network operations manually, they do recognize the advantages of automation.

Perceived advantages and disadvantages of automation

More than 65% of respondents indicated that the automation technologies will help them in tackling boring and monotonous jobs, or in saving money.

Very few indicated that automation would leave a lot of time for them to do other creative innovations, or that automation will be intelligent enough to revolutionize networking.

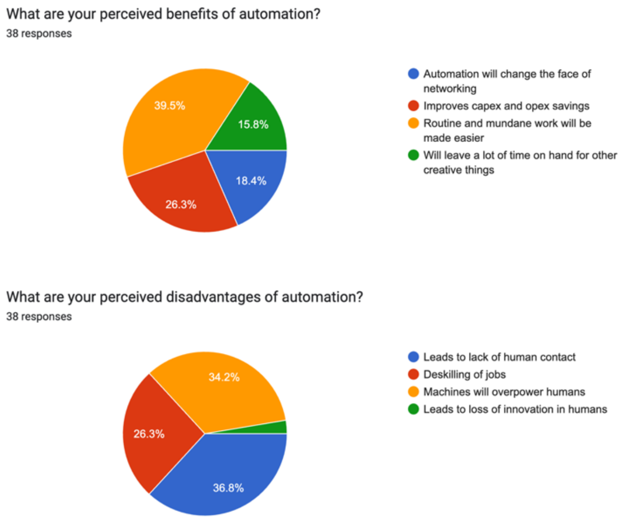


Fig 5 Advantages and disadvantages of automation

This indicates that the current set of automation suites, tools or approaches that are being considered are aimed to increase the ROI, and also slowly replace routine tasks with something that can be trusted to run automatically. Higher automation with intelligent tools meant for self-driven networks is something that is yet to be understood and acknowledged by the network operator community.

On the other hand, the lack of human contact was the biggest drawback of automation. The second-highest perceived disadvantage was that humans would lose control to machines, who will eventually outnumber them. Also, four out of six respondents who stated that automation results in employment deskilling were between the ages of 31 and 40, which may suggest their perception or worry of losing the jobs that their current skill sets are suitable for. **Fig 5** captures the responses indicating the advantages and disadvantages of automation.

Apprehensions and anxieties to adopt automation

While as noted above, most people agreed that moving to automation is beneficial, there was a mixed view on the apprehensions of moving operations from manual to automation. While most people (>70%) believed that they can transition to automation, there were still a few who believed that it would be very difficult to move to automation.

Further, the respondents also indicated that it was not quite easy to automate all of their network operations.

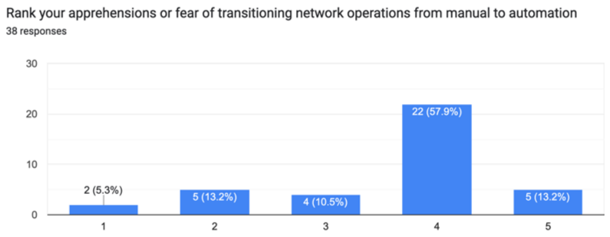


Fig 6 Apprehensions or fear of transitioning to automation

Majority believed that only 20%-30% of the operations could be automated, and none believed that they could cross 70% of their network in transitioning from manual to automated operations. **Fig 6** shows the responses ranking apprehensions in transitioning to automation.

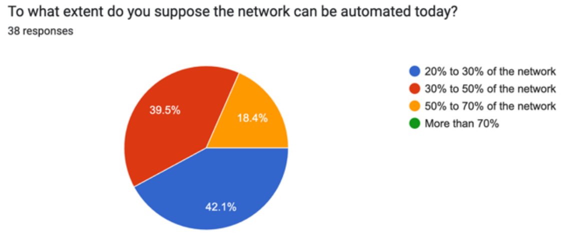


Fig 7 Extent of automating network operations

When observing the hurdles and problems associated with migrating to automation, the respondents reported that the initial challenge was a lack of understanding of automation technologies. Not enough trainings were offered to the existing staff either, nor was there enough budget to procure automation. **Fig 7** captures the responses indicating the extent by which network operations can be automated.

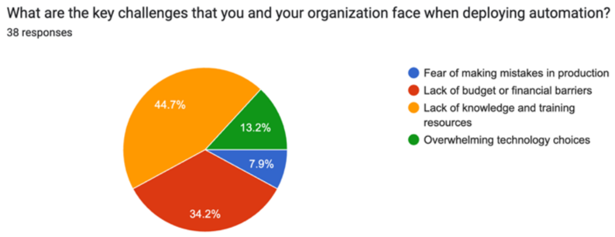


Fig 8 Key challenges faced when deploying automation

This basically means that even while the majority of firms think automation can be helpful to them, they lack the resources to buy a quality automation product and the necessary skill sets to experiment fully on their own. Another group of respondents claimed that the technology options were overwhelming, which was connected with their concerns that the move to automation would not be straightforward or painless. **Fig 8** shows the responses on key challenges faced while deploying automation.

Fear of network automation based on age factor

While operators in the age range of 30 to 50 looked to be either neutral or less comfortable in learning new technologies, respondents in the 18 to 30 age range expressed that they were either extremely comfortable or comfortable learning new automation tools and technologies. In a similar vein, only 2 respondents in the 31 to 50 age range described learning new technology as "fun," while the rest indicated that that they found the process "tiring" or "grueling." While it is very hard to come to a solid conclusion that all network technicians who were in the higher age bracket fear automation, it can be safely concluded that those who have been working manually on the network operations for a longer time need to be assisted with more confidence, motivation, and adequate training in order to transition them towards the changing requirements.

Results from customer interviews

Here are the salient points concluded from the principal network operator interviews:

Apprehensions in implementing vendor-based tools: One of the key concerns of network operators was to understand *who owns the knowledge of the network automation.* If the automation tool procured from the vendor breaks down, the challenges associated in getting the problem fixed seemed humongous. Another critical concern was also the lack of skillsets needed to understand the overwhelming technology evolution in the field of automation. If a large network adopting a complex automation technology breaks down, it would not only bring the network to a standstill, but also the lack of staff with proper skill sets may result in prolonging the downtime further.

In-house automation is preferred in such situations. However, in order to develop automation inhouse, adequate skill sets are required and with so many overwhelming technology concepts and automation languages, making the right choice is never easy.

Control is the key: Losing human control to self-managed and self-fixing automated networks is extremely daunting to network operators, especially if there is no way for them to take the control back in production networks. They are also concerned about deskilling of jobs by putting the trust in a machine, the functioning of which is unpredictable.

Single pane of glass and managing legacy networks: Network operators are concerned that all the upcoming technologies, be it SDN or CCLs require specific protocol and feature requirements that may not be available in legacy devices. Also, not all vendors pay attention to the standards, and in cases where the standards themselves are not fully mature, it is very hard to embrace the technology fully. On one hand few operators indicated that the IP networks and optical networks are typically operated by different vendors, and the lack of a single pane of glass made it very cumbersome to manage every aspect of the automation. On the other hand, few other operators expressed the fear that if a sole automation tool was used throughout the network, then the effect of the tool breakdown would be massive.

Return of investment (ROI) on implementing the automation technology: Automation boasts on improving the opex and therefore there is a need to demonstrate the time and costs saved by the technology performing the same job as humans, but in a much-optimized manner. Unless this ROI is demonstrated, it is very hard to replace the manual network operations with technology, as well as allocate sufficient funds to procure the technology.

IV. CONCLUSIONS

The purpose of this paper is to identify various challenges that network operators face in their NOCs and to comprehend why automation adoption is so slow in comparison with the growth of automation technologies.

SDN, CCLs, AI, and associated technologies have all developed to address various deployment types. Each have their own advantages as well as challenges and concerns. Adding to this, the nature of today's networks with multivendor devices, legacy equipment, non-adoption of standards and lack of a single orchestration platform is causing exhaustion to network operators. Although there is enough demand in automation to consider its adoption, it also seems that network technologies are developing rapidly. Network operators need to gain confidence and trust to use and accept them.

Vendor organizations will therefore have to find means to deliver automation to NOCs in such a way that it's flexible enough to implement in a modular form, rather than being forced to implement all at a time. Network operators are overwhelmed with technologies, and they are looking for simplistic and functional tool without causing any collateral damages. The ability to provide operators with an easy button to pause the automation in a fail-safe manner (such as providing a handle to change policy-sets, decision-logic modules and so on) is critical to gain the operator trust. When resumed, the automation framework must be able to detect the delta configurations that were performed manually in the network and synchronize the changes after obtaining the operator's approvals. Maintaining a log of actions and events, so that the operator or anyone else can retrace the activities being performed in the network by the automation tool, as well as maintaining a structured and authoritative source of truth for all the data that is collected for data monitoring and management is very crucial to improve the operator trust and thereby expedite the automation adoption.

Although an attempt has been made to conduct real-time interviews and to solicit insights from operators directly, this research has gathered data from a limited set of operators only. Also, although efforts were made to distribute the questionnaire to as many different types of people, there was a tad-bit concentration in the age group (mostly in the 18-30 age range).

Therefore, this research is to be taken as the initial steps to understanding the slowness in the adoption of network automation. A more detailed study on the concerns in the adoption of automation with deeper data collection and analysis is required for a further thorough understanding.

REFERENCES

1. Zhang, H., & Quan, W. (2021). Networking Automation and Intelligence: A New Era of Network Innovation. Engineering. https://doi.org/10.1016/j.eng.2021.06.019
2. Jefia, A., Popoola, S., & Atayero, A. (2018). Software-Defined Networking: Current Trends, Challenges, and Future Directions. http://ieomsociety.org/dc2018/papers/435.pdf
3. Kreutz, D., Ramos, F. M. V., Esteves Verissimo, P., Esteve Rothenberg, C., Azodolmolky, S., & Uhlig, S. (2015). Software-Defined Networking: A Comprehensive Survey. Proceedings of the IEEE, 103(1), 14–76. https://doi.org/10.1109/jproc.2014.2371999
4. Jammal, M., Singh, T., Shami, A., Asal, R., & Li, Y. (2014). Software defined networking: State of the art and research challenges. Computer Networks, 72, 74–98. https://doi.org/10.1016/j.comnet.2014.07.004
5. Jarraya, Y., Madi, T., & Debbabi, M. (2014). A Survey and a Layered Taxonomy of Software-Defined Networking. IEEE Communications Surveys & Tutorials, 16(4), 1955–1980. https://doi.org/10.1109/comst.2014.2320094
6. Eissa, H. A., Bozed, K. A., & Younis, H. (2019). Software Defined Networking. 2019 19th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA). https://www.academia.edu/9851597/Software_Defined_Networking
7. Ahmad, S., & Hussain Mir, A. (2022). Wireless and Microwave Technologies. 2, 11–32. https://doi.org/10.5815/ijwmt.2022.02.02
8. Vaishnavi, I., & Ciavaglia, L. (2020), November 1). Challenges Towards Automation of Live Telco Network Management: Closed Control Loops. IEEE Xplore. https://doi.org/10.23919/CNSM50824.2020.9269048
9. Boutaba, R., Shahriar, N., Salahuddin, M. A., Chowdhury, S. R., Saha, N., & James, A. (2021). AI-driven Closed-loop Automation in 5G and beyond Mobile Networks. Proceedings of the 4th FlexNets Workshop on Flexible Networks Artificial Intelligence Supported Network Flexibility and Agility. https://doi.org/10.1145/3472735.3474458
10. Ivanov, S., Kuyumdzhiev, M., & Webster, C. (2020). Automation fears: Drivers and solutions. Technology in Society, 63, 101431. https://doi.org/10.1016/j.techsoc.2020.101431

Spatio-Temporal Patterns and Explanatory Factors of Urban Fire Occurrences in New Taipei City

Ting-Jen Lo, and Yihjia Tsai*

***Abstract*— In modern metropolitan governance, fire-fighting represents a challenging task given the limited resources and the nature of disastrous impact on the citizen's psyche regarding the loss of both property and human lives. Fire-fighting and reduction of fire loss remain a challenging task for policy makers. The study of past fire incidences and factors empower policy makers for making resource allocation decision. In the wake of availability of large collections of both time-stamped and geocoded fire incident data, the exploration of any possible spatio-temporal patterns can help better allocation of limited fire resources. This paper employs spatio-temporal data analysis techniques such as spatial kernel density and spatial dependency index such as Ripley's K, Moran's I and Lee's L to explore socioeconomical factors and fire incidents. The results showed that in addition to population factors, the amount of electricity usage and ems demands are also relevant indicators for fire occurrences.**

***Index Terms*— fire incidence, spatio-temporal patterns, correlation.**

I. INTRODUCTION

Located in the northern part of Taiwan island, New Taipei City is a metropolitan with approximately 3.9 million population spreading in an area of over 2,000 km^2. The Fire Department of New Taipei City receives over 7,000 fire related calls each year reporting potential fire. Among those calls, in 2016, there are 161 severe fire incidences which took 27 human lives. The cause of fire is a chain of complex events, however, the occurrence of fire incidences exhibits particular spatial and temporal patterns. The availability of more detailed data reporting and data analytics techniques enable the development of predictive models to access various risks of fire, and prioritize fire inspection [8]. The use of spatio-temporal correlation analysis of fire occurrence patterns has been applied mostly to forest or wild fires, recent years there are several studies on urban residential fire [9]. In a densely populated city area, residential fires pose a threat to human lives much more than forest fire. An investigation of the patterns of fire and risk factors could lead to identification of fire hazards and creation of measures to reduce the fire threat in the city.

II. RELATED WORK

In the past, most research efforts related to fire occurrences have focused on wild fire, however, the recent decade have seen a growing body of literature in urban fire studies. The expanding influence of big data analytics has gradually attracted researchers to the area of urban fire. Factors associated with urban fire is a complex mixture of indicators involving human activities, housing locations, and structures [10], for example, vacant and abandoned buildings, declining neighborhoods, housing ownerships, household income levels, educational achievements, ethnicity, age groups, smoking, alcohol and drug abuse and so on. Some study even suggests that the degree of association between those indicators and fire incidents are context sensitive, as in the case of educational level. [12]. In addition to socioeconomic factors, the association between urban fire and weather conditions, calendar events are examined in [3]. Their findings suggest that the risk of fire increases when school holidays and during long weekends. Also, high temperatures and neighborhoods of lower socioeconomic indexes are prone to increased fire risks.

Four years of fire incidence data (from 2000/01/01 to 2004/12/31) were used to study the spatial distribution of urban fire incidence [2]. The data include four fire incident types: FDR1 (fire involving property), FDR1V (vehicles), FDR3 (derelict buildings/vehicles, outdoor structures), and FAM (false alarm deemed malicious). The spatial distribution of overall incident rate (per 1,000 population) is used in comparing each census wards. In addition to census wards comparison, KDE(kernel density estimate) for point distributions are calculated based on empirically derived bandwidth parameter of 2 km. The authors in [2] also explores the relation between fire incident counts and 32 explanatory variables. Seven natural groups are used to organize those variables: car ownership, educational attainment, ethnicity, family structure, housing ownership, age profile, and household structure.

The results in [2] indicated that FDR1 type of fire incidents are more prone to occur in wards with lower level of educational attainments as well as lower proportions of white residents. Lower educational attainment are associated with FDR1V and FDR3 fire incidents while lower proportions of childless couples and lower proportions of car owners are more prone to FAM (false alarm) fire incidents.

All authors are with the Department of Computer Science and Information Engineering, Tamkang University, Taipei, Taiwan, R.O.C. (email: roba641068@gmail.com, iaiclab.tku@gmail.com).

The association between urban fire and different levels of urbanization are reported in [11]. The data used in the study are monthly distribution of fire incidents from 2006 to 2008 in two cities: Beijing, and Jinan, and the city of Hefei has only monthly data from 2007 to 2008. The three cities representing different levels of urbanization features. Beijing is a highly urbanized city while Jinan and Hefei are both moderately developing cities. Their findings indicates that the more urbanized a city is, the severity of fire increases and the cause of fire are majorly attributed to electricity and careless use of fire. One temporal pattern discovered from the study is that the first quarter of the year receives more fire occurrences than the other three quarters.

A study of 12-year (2002-2013) fire event data in Nanjing city indicates that dwelling (DW) and facility (FL) fires are the majority type of urban fire [13]. Kernel density estimate technique is used to calculate the spatial hot spots distribution, and the empirical bandwidth used is 1.0 km. A 1.0 km x 1.0 km grid is used to divide the city into rectangular areas and the local Moran's I index are calculated for each square to find high incident area surrounded by high fire incident counts squares.

Temporal variations of different time scales are also reported in [13]. Their results indicate that monthly distribution of DW type of fire is generally higher than FL type of fire and DW fire peaks at September while FL fire peaks at July and August. As for weekly time scale, DW type of fire is more likely to occur on Sunday while most FL fire events occurred on Tuesday. For hourly time scale, most of DW fire appeared during 8AM to 9PM while a large portion of FL fire occurred during 9AM to 11PM. Univariate and bivariate co-maps are used in the report to study the spatial-temporal patterns of fire incidences [13], however, as the authors noted, different definition of subset in the co-map results in different patterns. Thus, the issue of choosing a proper subset to be included in the co-map is yet to be investigated. Several summary statistics are commonly used in determining whether a spatial point distribution occurred randomly in space. Ripley's *K* and *L* functions are well known summary statistics to test against random Poisson point process assumption. Those two functions are closely related statistics for spatial point pattern analysis, they are used to examine the amount of deviations from spatial homogeneity. The formula for calculating sample-based estimates of Ripley's *K* and *L* functions are:

$$\widehat{K}(r) = \frac{1}{\lambda}\sum_{i\neq j}\frac{\mathbb{I}[d_{ij}<r]}{n} \qquad (1)$$

$$\widehat{L}(r) = \left(\frac{\widehat{K}(r)}{\pi}\right)^{1/2} \qquad (2)$$

where λ is the average density of all points, r is the search radius, and d_{ij} is the distance between point i and j in a set of n points. $\mathbb{I}[d_{ij} < r]$ is the indicator function that returns 1 if the distance between points i and j, d_{ij}, is less than r and returns 0 if otherwise. If the spatial distribution of the points are approximately homogeneous, then the value of $\widehat{K}(r)$ should be approximately equal to πr^2.

One practice is to draw the graph of $\widehat{L}(r) - r$ against r instead of $\widehat{L}(r)$. The purpose of this is to better differentiate the deviation at low r values. If the data follow a homogeneous Poisson process, then the plot will approximately follow the horizontal zero axis with random dispersion. The Ripley's *K* function allows one to determine whether the data points follow a random, dispersed or clustered distribution pattern at certain scale. Besides the commonly used Ripley's *K*-function for spatial point pattern analysis, the authors in [1] proposed the use of Diggle's *D*-function for spatial clustering detection of one distribution against another underlying non-Poisson distribution and spatio-temporal *K*-function for space-time interaction analysis. In this paper, a similar approach will be used to determine whether the spatial distribution of fire incidents deviates from Poisson distributed point process within a census tract.

III. DATA DESCRIPTION AND METHOD

From census data dated July 2018, the total area of New Taipei city is 2052.57 km^2 and the population is 3,988,906. New Taipei city is composed of 29 districts that are further divided into 1,032 villages, which is the smallest census unit at the moment. Population density of those 1,032 villages varied from 5 person per square kilometers to 119,679 person per square kilometers while the area of villages ranged from 0.0167 to 183.87 km^2.

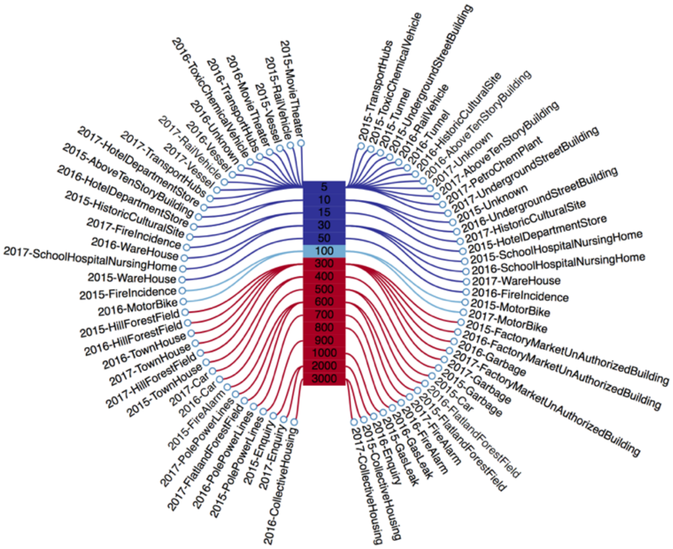


Fig. 1 Yearly distribution of fire incidence types (2015-2017)

From 2015 to 2017, the Fire Department of New Taipei City receives more than 7,000 fire incidents yearly and those incidents were categorized into 23 incident types according to the nature of fire. From 2015 to 2017 the number of incidents for each incident type are plotted in Fig. 1, the most frequent fire incident type is "collective housing" in all three year's statistics. From 2015 to 2017, the percentage of "collective housing" type of fire incident takes up 27.67%, 31.51%, and 33.61% respectively. The second frequent fire incident type is “Enquiry” for all three years. This type of fire incident represents minor fire incidents that could serve as indication of possible fire if not properly taken care of. In both years 2015 and 2016 the “gas leak” type was the third most frequent fire incident, however, in year 2017 it was changed to became part of “Enquiry” type because natural gas leaking is not necessarily resulting in fire incident.

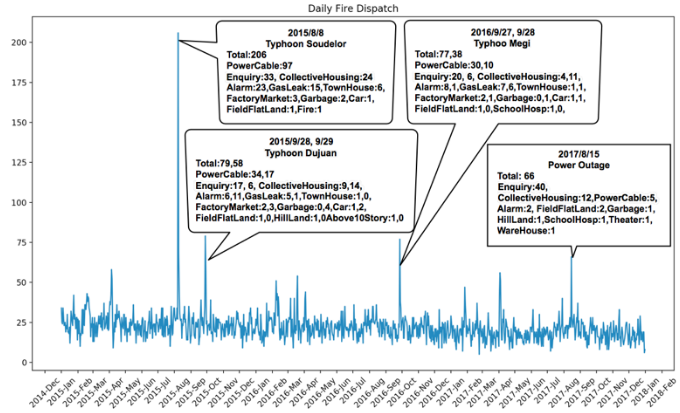


Fig. 2 Daily fire counts (2015-2017)

A. Temporal Patterns

Fig. 2 plots the daily occurrences of fire incidents from year 2015 to 2017, the most prominent feature of the series is that there is an unusually high peak in August 2015. Further investigation of those high occurrences of daily fire reveals that all incidences more than 70 are caused by tropical cyclone, also known as 'typhoon'. Each summer season, Taiwan is at the forefront of south-pacific typhoon. A super typhoon came with high winds and heavy rain that may cause drastic damages. As illustrated in Figure, there are two events in year 2015 and one event in 2016 that are caused by Typhoon, however, there is no typhoon hits the island in 2017, and the only peak is caused by power outage.

By averaging the daily fire counts for each month and look at each year separately, three radar diagrams are plotted using the average daily fire incident counts for the 12 months, the results are shown in Fig. 3.

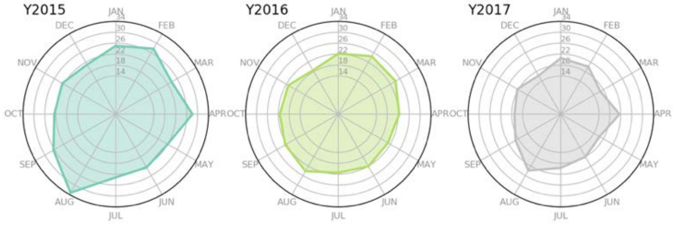


Fig. 3 Avg. of daily fire counts in each month (2015-2017)

The three radar plots representing each year showed a consistent tendency of three peaks in the August, February and April. Those are caused by the special event effects, the lunar new year's period in the February, the tomb sweeping holiday in early April and typhoon season during August.

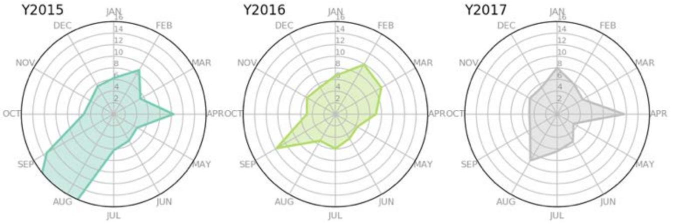


Fig. 4 Std. dev. of daily fire counts in each month (2015-2017)

From the standard deviation of daily fire incident counts in Fig. 4, the largest variance appears in around August and September if there is typhoon landed in New Taipei city. The second peak of variance located in February and April. These are consistent with the sequence plot in Fig. 2. At a finer time scale, the fire incident data reveals another pattern that would be consistent with findings in other related research. As shown in Fig. 5, the average hourly data for different day of week have shown a declination from 23:00 till 09:00, similar results were also being observed in [13]. The average amount of fire incident is smallest when in Wednesday and Thursday, and during weekend and Monday, the average amount of fire incident become larger than other day of week. Those findings are consistent with literature where long weekend and holiday are prone to fire risks [3].

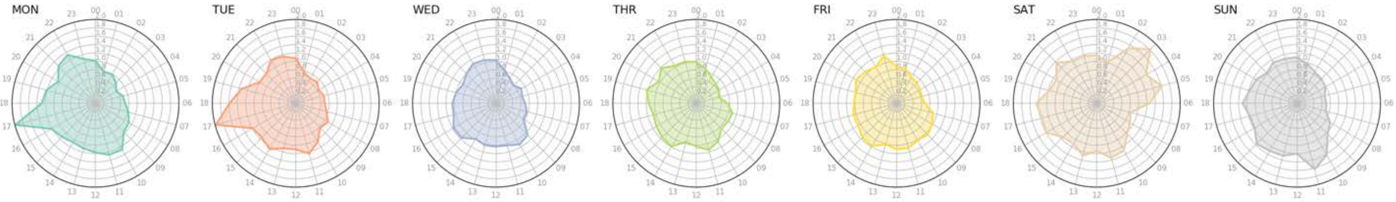


Fig. 5 Avg. of hourly fire counts by day of week (2015-2017)

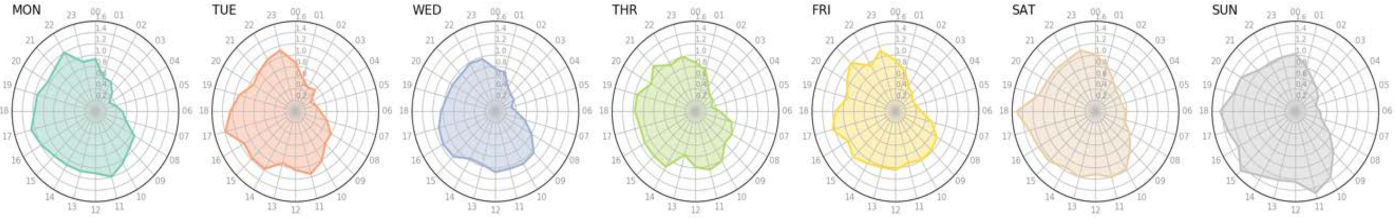


Fig. 6 Std. dev. of hourly fire counts by day of week (2015-2017)

In Fig. 6, the std. dev. of the hourly data for different day of week showed that variations of fire incidents in Saturday is larger than all other weekdays.

B. Spatial Patterns

The spatial distribution of yearly fire incident counts among all 1,032 villages are illustrated in Fig. 7. Visually comparing the three plots in Figure, one can conclude immediately there are similar clusters between those three plots. All dark blocks are located on the western portion of New Taipei city. White blocks indicating no reported fire incidents during that year are located on the south eastern portion and northern part of New Taipei city. These areas are sparse in population due to high mountains.

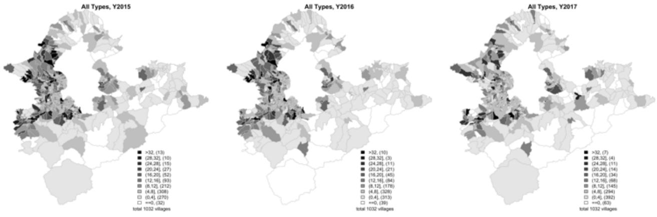


Fig. 7 Spatial distributions of yearly fire counts (2015-2017)

While the total amount of fire incident counts in each village are used to create Fig. 7, and the concentration of population may have a direct consequence of more fire incidents reported. It would be natural to explore the adjusted choropleth maps in which fire incident counts are divided by thousand population counts. The resulting three year's population adjusted fire density plots are shown in Fig. 8.

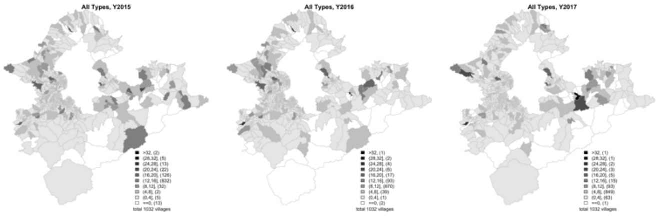


Fig. 8 Spatial distributions of population adjusted yearly fire counts (2015-2017)

Fig. 8 illustrated a different scenario than in Fig. 7, the darkest blocks are not as clustered as those in the corresponding unadjusted map. The choropleth maps of population adjusted fire incidents showed those areas where population cannot directly explaining the amount of fire incidents.

In the analysis of spatial point distributions, statistics used to measure the degree of spatial homogeneity would be the Ripley's *K* function and it's close relative Ripley's *L* function, as shown in Equation (1) and Equation (2). Note that Ripley's *K* and *L* functions are summary statistics, they compare the observed pattern of data with that generated by a homogeneous Poisson process. That is, those two statistics are indications of degree of deviation from the complete spatial randomness (CSR) presumption. They do not give exact location of where clusters might appear. Also, when the density of samples are not uniformly distributed among the spatial domain, the statistics may give out the results of clustered data points without any adjustment for the different underlying spatial density. Other summary statistics that belonging to this family are *G/F/J*-functions. The *G* function describes the cumulative distribution of the nearest-neighbor distances for a typical point in the data set. The empirical estimate of *G* function for a set of *n* data points is formulated as:

$$\hat{G}(r) = \frac{1}{n}\sum_{i=1}^{n} b_g(i,r)\mathbb{I}[d_{ij} < r] \qquad (3)$$

where $b_g(i,r)$ is an edge correction factor so that $\hat{G}(r)$ would be approximately unbiased. The *F* function is a measure of cumulative distribution of empty space distance. Suppose s_u represents the empty space distance from any fix location *u*, that is, $s_u = \min_{\forall i} d_{ui}$ for all points *i* in the data set. Empirical estimate of *F* function based on a grid of locations *j*, $j = 1,2,\cdots,m$, is given by

$$\hat{F}(r) = \frac{1}{m}\sum_{j=1}^{m} b_f(j,r)\mathbb{I}[s_j < r] \qquad (4)$$

where $b_f(j,r)$ is an edge correction weight for location *j*. For comparison, the homogeneous Poisson point process of intensity λ for both *G* and *F* functions is given by

$$G_{\text{pois}}(r) = F_{\text{pois}}(r) = 1 - \exp(-\lambda\pi r^2) \qquad (5)$$

The *J* function is a combination of *F* and *G* functions, it is formulated as:

$$J(r) = \frac{1-G(r)}{1-F(r)} \qquad (6)$$

When a point pattern is known to be spatially inhomogeneous, a modification of the *K* function to account for the inhomogeneity is given as

$$\hat{K}_{\text{inhom}}(r) = \frac{1}{|A|}\sum_{i=1}^{n}\sum_{j\neq i}^{n} b_k(i,j,r)\frac{\mathbb{I}[d_{ij}<r]}{\hat{\lambda}(i)\hat{\lambda}(j)} \qquad (7)$$

where |A| represents the area of the region while $\hat{\lambda}(i)$ is an estimate of the intensity function *λ(i)* at point location *i*. The edge correction weight function is denoted by $b_k(i,j,r)$. In addition to summary statistics, most studies of single variable spatial patterns subscribe to the use of Moran's I index and the local version of the Moran's I index. Local Moran's I is given by

$$\hat{G}(r) = \frac{1}{n}\sum_{i=1}^{n} b_g(i,r)\mathbb{I}[d_{ij} < r] \qquad (8)$$

Widely used tests for space-time interaction with spatiotemporal point processes include Diggle et al. [4], Knox [5], and Mantel [6] indicators. Diggle et al. [4] extended Ripley's *K* function to include spatiotemporal version, the formula can be written as:

$$\hat{K}(r,t) = \frac{1}{\hat{\lambda}^2|A\times T|}\sum_{i=1}^{n}\sum_{j\neq i}^{n} \mathbb{I}[d_{ij} < r]\mathbb{I}[t_{ij} < t] \qquad (9)$$

IV. RESULTS

In this section, tests for spatial clustering of three years of fire incident data are reported. First, the results of kernel density estimates of three year's fire incident data are illustrated in Figure. The bandwidth chosen is 1km and all three plots demonstrated similar distributions of hotspots. The distributions of hotspots in those three plots are indicative of spatial population density clustering.

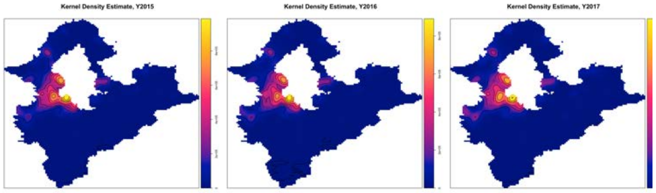

Kernel Density Estimate, Y2015
Kernel Density Estimate, Y2016
Kernel Density Estimate, Y2017

Fig. 11 Kernel Density Estimates of yearly fire incident counts (2015-2017)

By aggregate three years of fire incident counts and calculate the overall summary statistics, the results are shown in Figure. Comparison of the empirical summary statistics and the theoretical homogeneous Poisson point processes, all four statistics indicate spatial clustering.

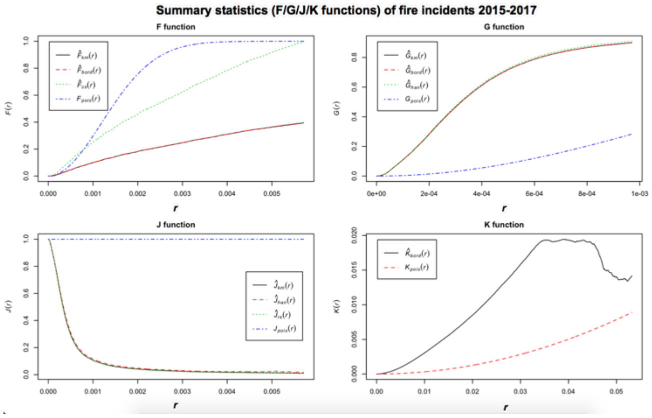


Fig. 12 Summary Statistics of fire incident counts (2015-2017)

Please note that these four summary statistics are all comparing the data set against the distribution of homogeneous Poisson point process and the results showed spatial clustering. As the kernel density estimate plots illustrated in Figure, it is not surprising from those four summary statistics. Further investigation can be done by the use of inhomogeneous *K* function.

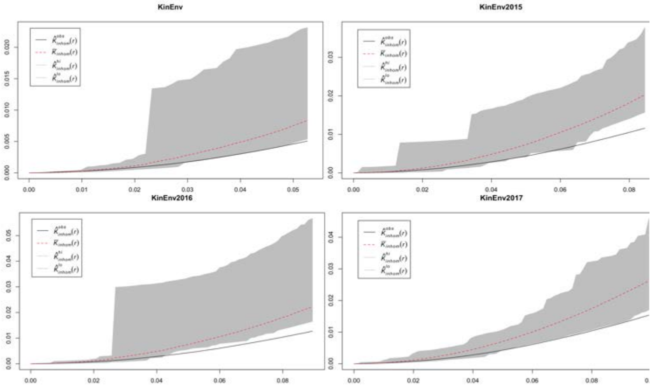


Fig. 13 Inhomogeneous K functions for fire counts (2015-2017)

After adjusting for inhomogeneity, the results in Figure showed a inhibition pattern than suggested.

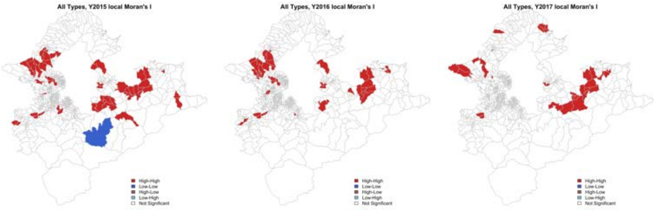

All Types, Y2015 local Moran's I
All Types, Y2016 local Moran's I
All Types, Y2017 local Moran's I
High-High
Low-Low
High-Low
Low-High
Not Significant

Fig. 14 Local Moran's I of population adjusted yearly fire counts (2015-2017)

Using population-related data from the government open-data archive as covariate, the results of correlations between variables are shown in Figure. Those census tract data are number of hydrants in the village, education level above graduate schools, above university level, an Fire2017(fire incident counts).

Some socio-economic statistics are gathered from both the open-data archive and the Fire Department of New Taipei City, for example, the number and dispatch type of calls for emergency medical services in each village.

V. CONCLUSIONS

This paper analyzes three years of fire incident data for New Taipei city from 2015 to 2017, in addition to temporal and spatial patterns, publicly available contemporaneous open-data are used as covariates to explanatory indicators for risk of fire. Findings are that urban fire incidents subject to severe weather conditions such as typhoon and there is a tendency of increased fire risk for weekends as well as traditional holidays. Vacant houses are clearly an indication of low fire incident counts, contrary to what have been reported in the researches in the US [7].

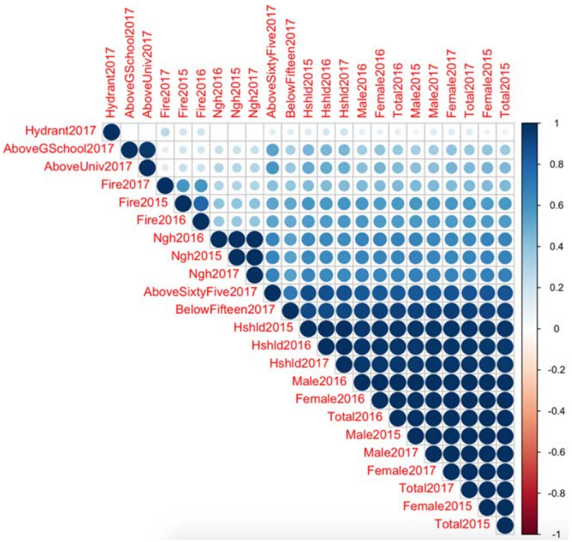


Fig. 15 Correlation between fire incidences and population-related factors

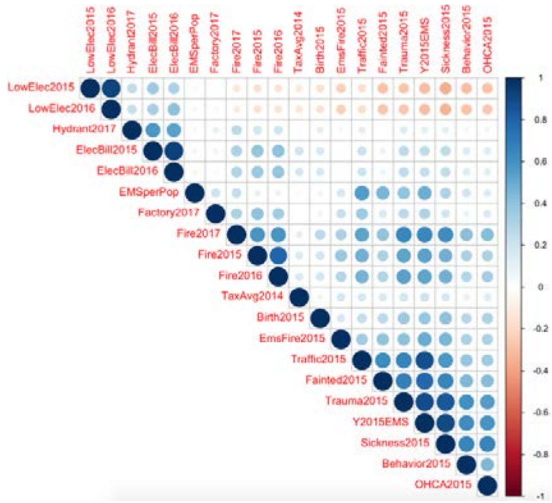


Fig. 16 Correlation between fire incidences and socioeconomic-related factors

References

[1] E. Ceyhan, K. Ertuğay, and Ş. Düzgün, “Exploratory and inferential methods for spatio- temporal analysis of residential fire clustering in urban areas,” Fire Safety Journal , vol. 58, pp. 226–239, 2013.

[2] J. Corcoran, G. Higgs, C. Brunsdon, A. Ware, and P. Norman, “The use of spatial analytical techniques to explore patterns of fire incidence: A South Wales case study,” Computers, Environment and Urban Systems, vol. 31, no. 6, pp. 623–647, 2007.

[3] J. Corcoran, G. Higgs, D. Rohde, and P. Chhetri, “Investigating the association between weather conditions, calendar events and socio-economic patterns with trends in fire incidence: An Australian case study,” Journal of Geographical Systems, vol. 13, no. 2, pp. 193–226, 2011.

[4] P. J. Diggle, A. G. Chetwynd, R. Häggkvist, and S. E. Morris, “Second-order analysis of space-time clustering,” Statistical Methods in Medical Research, vol. 4, no. 2, pp. 124–136, 1995.

[5] E. G. Knox and M. S. Bartlett, “The detection of space-time interactions,” Journal of the Royal Statistical Society. Series C (Applied Statistics), vol. 13, no. 1, pp. 25–30, 1964.

[6] N. Mantel, “The detection of disease clustering and a generalized regression approach,” Cancer research, vol. 27, no. 2 Part 1, pp. 209–220, 1967.

[7] National Fire Data Center, “Vacant residential building fires (2013-2015),” Topical Fire Report Series, vol. 18, no. 9, p. 12, 2018.

[8] J. Roman, “In pursuit of SMART,” NFPA Journal, vol. November/December, pp. 41–50, 2014.

[9] S. Strydom and M. J. Savage, “A spatio-temporal analysis of fires in South Africa,” South African Journal of Sciences, vol. 112, no. 11/12, p. 8, 2016.

[10] United States Fire Administration, “Socioeconomic factors and the incidence of fire,” Federal Emergency Management Agency, Report FA 170, 1997.

[11] J. H. Wang, J. H. Sun, S. M. Lo, L. Gao, and R. Yuan, “Statistical analysis on the temporal- spatial characteristics of urban fires under typical urbanization features,” Procedia Engineering, vol. 11, pp. 437–444, 2011.

[12] L. Yang, Y. Yang, W. Cui, J. Gong, and T. Fang, “The relationships between socioeconomic factors and fire in China,” in Proceedings of the Asia- Oceania Symposium on Fire Science & Technology (AOFST 6), 2004, p. 6.

[13] J. Yao and X. Zhang, “Spatial-temporal dynamics of urban fire incidents: A case study of Nan-jing, China,” in The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (2016 XXIII ISPRS Congress), (Prague, Czech Republic), vol. XLI-B2, 2016, pp. 63–69.