Advances in Artificial Intelligence and Machine Learning; Research 2 (3) 441-455 Received 10-07-22; Accepted 26-07-22; Published 01-08-22

AI Based Approach for Shop Classification and a Comparative Study with Human

Mrouj Almuhajri mr_ almuh@encs.concordia.ca
Centre for Pattern Recognition and Machine Intelligence
Concordia University, Montreal, QC, H3G 2W1, CA
Department of Computer Science Saudi Electronic University, KSA

Ching Y. Suen suen@cse.concordia.ca
Centre for Pattern Recognition and Machine Intelligence
Concordia University, Montreal, QC, H3G 2W1, CA

Corresponding Author: Mrouj Almuhajri

Copyright © 2022 Mrouj Almuhajri and Ching Y. Suen This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

The rapid advancements in artificial intelligence algorithms have sharpened the focus on street signs due to their prevalence. Some street signs have consistent shapes and pre-defined colors and fonts, such as traffic signs while others are characterized by their visual variability like shop signboards. This variations create a complicated challenge for AI-based systems to classify them. In this paper, the annotation of the ShoS dataset were extended to include more attributes for shop classification. Then, two classifiers were trained and tested utilizing the extended ShoS dataset. SVM showed great performance as its F1-score reached 89.33%. The classification performance was compared with human performance, and the results showed that our classifier excelled over human performance by about 15%. The results were discussed, so the factors that affect classification were provided for further enhancement.

Keywords: Shop classification, ShoS dataset, SVM, NLP, Signboard

1. INTRODUCTION

Advancements in artificial intelligence algorithms have been extended to street signs due to their prevalence. Nowadays, complex algorithms can be executed in real-time on contemporary mobile devices to process our surrounding environment. While Some street signs have standardized shapes and pre-defined colors and fonts, such as traffic signs, others are characterized by their visual variability like shop signboards. These variations have presented a complicated challenge to AI-based systems to detect and classify them. Store signboards represent about 41.5% of other street sign types which increases the demand to build systems that can identify them [1]. This research was driven by possible beneficial applications for humans, municipal agencies, and automobiles.

Citation: Mrouj Almuhajri and Ching Y. Suen AI Based Approach for Shop Classification and a Comparative Study with Human. Advances in Artificial Intelligence and Machine Learning. 2022;2(3):29.

Different types of users would benefit from these applications, such as tourists and visually impaired individuals. According to the World Health Organization[1], there are 2.2 billion people at least around the world living with vision impairment or blindness. Such AI-based systems can improve the quality of their lives by leveraging smart phones to explore new neighborhoods and recognize any store. This can be done simply by taking a picture of storefronts regardless of how they see it or in which language the signboard is without the need for human assistance.

In addition, municipal government agencies develop norms and regulations governing the appearance of shops and signboards. Human inspectors are typically utilized to ensure that such regulations are followed. However, this can be time consuming and prone to human errors. Hence, a system for identifying and classifying shop signboards can improve the speed and accuracy of this process. For instance, the City of Westmount in Quebec issued regulations governing the design of storefronts and their signboards to mandate their harmony with the architectural design [2]. Language, size, lettering, and graphic components are covered in these guidelines.

Furthermore, even though self-driving vehicles are already utilizing AI-based systems in many aspects, there are still some other areas that can benefit from such systems. For example, some local stores are not registered on maps, or their information is not updated due to the rapid change in business openings and closings. Therefore, self-driving cars that are equipped with multiple cameras may utilize such systems to provide a more comprehensive analysis of their surroundings.

Although recognizing text in natural scenes has been an ongoing research topic for decades, many studies focused on the consistent appearance of text like in traffic signs [3], and license plates [4]. However, other signs like shop signboards have gotten less attention because of their complex visual appearance. We focused in this work on the text of store signboards as it contains rich semantics related to the store type.

This work contributes by 1) extend the annotation of the ShoS dataset [5], to include more attributes for shop classification purposes , 2) Train and test two classifiers using the extended version of the ShoS based on our previous work in [6], and 3) conduct a comparative study with human to measure the performance of our classifier versus human. The results were presented and discussed to show the factors that affected classification for further exploration.

2. RELATED WORK

2.1 Storefronts Detection and Classification

In [7], the proposed model detected the whole storefronts given street panoramic views using Multi-Box model uses a CNN. Although a large-scale dataset was used [8], it was difficult to annotate the huge volum of the dataset. So, a smaller amount of data was used in testing. The proposed method got a recall=91% compared to 62% for selective search while MCH map could not detect the boundary of storefronts precisely. This was because the fact that storefronts are more exposed to noise and they can abut each other.

[1] https://www.who.int/en/news-room/fact-sheets/detail/blindness-and-visual-impairment

In study [9], a system composed of several models was proposed for detecting the whole storefront. The detector was based on YOLOv3 [10], which detected the storefront and fed it to a classifier. The classifier extracts morphological and textual information and uses them as cues. The work was evaluated using a very limited dataset, and their methodology got mAP@0.50 = 79.37% for detection and 80.44% for classification.

In our previous work [5], the purpose was to detect signboards from street storefront views. The detectors were based on YOLOv3 and SSD. The models were trained and tested using the ShoS dataset over different variances of color schemes and input resolutions. The results showed that the YOLOv3 performed better without any color alteration. The mean average precision mAP@0.5 = 94.23% and 90.4% for YOLOv3 and SSD respectively.

In [8], a CNN model was trained to multi-label storefronts. Since the annotation was done by operators, some inconsistencies in the labelling were noticed which confused the learner. So, the authors built the model using ontological classifications in which a group of labels belong to a high-level class. During training, textual information, extracted by an OCR, was used beside the geo-information to match the extracted batches with the ground truth. In testing, the top k predictions were compared with the ground truth. An accuracy of 83% was achieved. The model results were compared with human performance by conducting two surveys. Their model performance found to be close enough to human-level accuracy.

(a) (b) (c)

Figure 1: Samples from the literature work for (a) storefront detection [7], (b) storefront detection and classification [9], and (c) store multi-label classification [8]

Most of the previous mentioned works used the whole storefront in their systems (see FIGURE 1), and faced some crucial issues in detection and classification because of 1) the limitation of existing datasets with full annotation, 2) the boundaries of storefronts are not clear enough to be learned, and 3) some irrelevant information that can be found in the storefront may mislead the classification process. That motivated us to tackle such issues by focusing on the text appearing in store signboards as it includes semantic textual information useful for classification.

2.2 Storefront Datasets

Many well-known datasets are available for text analysis in natural scenes. For example ICDAR 2013 [11], ICDAR 2015 [12], COCO-Text [13], and CTW [14], datasets are known for text localization and recognition purposes. The focus of these datasets is on detecting or recognizing the text itself. Hence, the images were not annotated for store classification.

Table 1: Summary of the storefront datasets compared to the ShoS dataset.

	SVT DS [15]	Google DS [8]	ShopSign DS [16]	ShoS DS [5,6] (ours)
Type	Public	Private	Public	Public
Source	GSV	GSV	Real SV	GSV
How collected	Amazon Mechanical Turk	Locally from Google data	Manually	Upwork freelance platform
Where	unknown	Europe, Australia, Americas	China	United States, Canada
Images	350	1.3m	25,770	10k signs within 7500 image
Annotation	Alex Sorokin's Toolkit	Operators	Manually	VGA Image Annotator+ operators
What	Words in nearby businesses	Whole storefront + labels	Characters in store signboards	Store signboards + store classes + words + typeface class + difficulty level
Classes	~50 lexicon words per image	208 store labels	4,072 unique Chinese characters	7 store classes 6 typeface classes
Language	Latin	Latin	Chinese mainly	English

Despite the large number of street view images provided by Google, there are still insufficient datasets provided for the field of detecting and classifying shops. Street View Text SVT dataset is one of the earliest public datasets that focused on the signage of retail business [15]. However, the annotation of the SVT only records bounding boxes around words in their signboards. In addition, Google dataset [8], was collected with Google collaboration to multi-label stores, but it is not provided to the public unfortunately. Moreover, the ShopSign dataset [16], was collected and annotated manually from real street views to capture Chinese shop sign images. Although it is large in scale, it is based on Chinese characters with limited images containing Latin characters. Finally, the ShoS dataset (ours) [5] is a public dataset collected from Google Street Views GSV to detect shop signboards. Its annotation is extended to include more attributes for classification purposes and it will be detailed more in Section 3. TABLE 1 summarizes the most comparable datasets to ours.

2.3 Text Classification Using NLP

Categorizing a group of documents into pre-defined classes is called supervised classification [17]. The field of text classification in natural language processing NLP is one of the well-studied areas for decades using machine learning. The era of deep learning added more advancement to the field, however in this research we applied some of the traditional machine learning techniques and avoid deep nets for the following reasons [18]: 1) text classification based on deep learning algorithms requires a tremendous amount of data for training compared with traditional methods, 2) the computational cost for deep learning approaches might get complicated during training and, 3) the limitation of a comprehensive theoretical understanding of the learning process increases, because of the concept of ”blackbox” nature in deep learning methods.

A comparative study [19], applied some of the text classification methods to classify news. Naive Bayes NB, K-nearest neighbor KNN, and SVM were tested on data collected from public news. Results showed that KNN and SVM performed better while NB was in the average range. Another work [20], compared Multinomial NB, KNN, and Decision Tree DT for topic classification. Data were collected from Amazon’s product reviews. The outcomes disclosed the superiority of MNB among the others with an F1-score reaching 91.8%. Moreover, Recurrent Neural Network RNN, SVM, and MNB were comparatively studied to classify spam in emails [21], with data acquired from Kaggle. In terms of F1-score, SVM was the best followed by MNB with 94% and 85% respectively. Considering the previous works, it is clear that traditional methods still perform better in some of the text classification problems.

3. DATASET

The Shop Signboard Dataset ShoS was used in this study as a base [5]. More attributes were added, as described in Section 3.1, to reach the targeted goal of classifying retails. The ShoS dataset contains 10k store signboards within 7500 storefront images. They were collected from Google Street Views GSV by some freelancers from Up Work platform[2]. The GSV images were captured from multiple Urban and suburb cities in Canada and the USA. An image can have a minimum of one storefront and up to seven storefronts. The shop signboards were cropped out of the full street view generating another dataset named **the ShoS-cropped** to enable usage of both datasets for multiple research purposes. FIGURE 2 shows some sample images from both datasets.

Figure 2: Some sample images from the ShoS dataset (left) and the ShoS-cropped dataset (right)

[2] https://www.upwork.com/

3.1 Data Annotation

The ShoS dataset was originally annotated for detection purposes using VIA tool [22]. The annotation included top-left and bottom-right coordinates plus the width and height of each signboard [5]. For classification purposes, the following attributes were added for each instance: sign text, store class, local or chain, occluded or not, and difficulty level. FIGURE 3 shows some samples from the CSV annotation file.

filename	image width	image height	region count	region id	xmin	ymin	width	height	xmax	ymax	sign text	class	occluded	city	local	difficulty
img403.png	1280	823	2	0	469	242	211	114	680	356	Pitt Bull	fashion	no	los angeles	1	2
img403.png	1280	823	2	1	776	249	176	127	952	376	CAPBANKS	fashion	no	los angeles	1	0
img404.png	1280	823	1	0	660	229	450	69	1110	298	JIO LUGGAGE J.J.J	other	no	los angeles	1	0
img405.png	3360	2100	2	0	831	580	966	251	1797	831	Great Lakes Pizza	rest_drink	no	sudbury	1	0
img405.png	3360	2100	2	1	1918	620	1175	206	3093	826	Cosmo Prof	health_pcare	no	sudbury	1	2
img406.png	1280	823	1	0	810	229	84	90	894	319	Jack in the box	rest_drink	no	los angeles	0	2

Figure 3: Some samples from the CSV annotation file after extending the attributes of the ShoS

The “sign text” attribute records the text that appeared on the signboard excluding numbers and addresses as they do not reflect essential information for store classification. The “class” attribute represents shop main classes, and they are based on the North American Industry Classification System NAICS. The scope of this study considered six main classes defined as the following:

1. **Restaurants and Drinking** which includes full/limited services restaurants, fast food restaurants, coffee houses, and bars.
2. **Food and Beverages** which includes grocery stores (supermarkets and convenience stores), and specialty food stores (meat markets, fish markets, and fruit/vegetables markets).
3. **Health and Personal Care** which includes clinics, pharmacies and drug stores, optical good stores, cosmetic/beauty supplies and perfume stores, and GYM.
4. **Finance and Investing** which includes banks, insurance, money marts, accounting and tax services, and real estate.
5. **Technology** which includes computer/electronic stores, telecommunication stores, and digital printing and copying services.
6. **Fashion** which includes clothing stores, jewelry/accessories stores, and shoe stores.

Any remaining categories go under the “Other” class. FIGURE 4 illustrates the hierarchy of the main and sub classes considered in this study.

For the “occluded” attribute, it could have two values (yes or no) where it represents if at least 20% of the text on the sign is covered by trees, traffic signs, big vehicles, spotlights, shadows, or other pedestrian signs. Moreover, the “local” attribute checks whether the store is a chain store (value=0) or a local store (value=1). This is to ensure that the ShoS dataset is not biased towards chain stores. Evidently, 84% of the ShoS stores are local. FIGURE 5 illustrates the ratio of local versus chain stores in general and for each class in the ShoS dataset. It is clear that class Finance and Investment

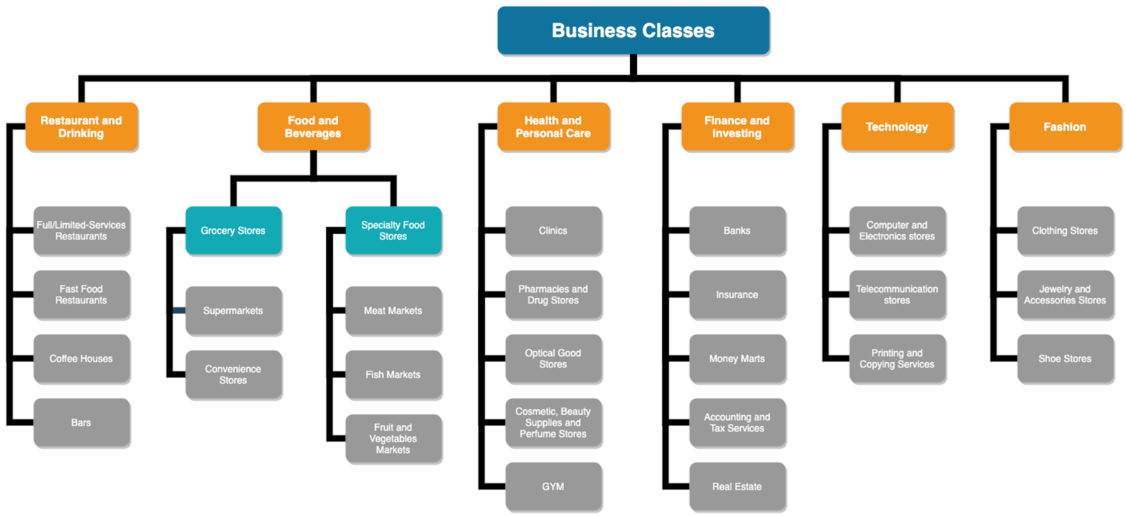


Figure 4: The hierarchy of the main and sub classes of shops considered in this study

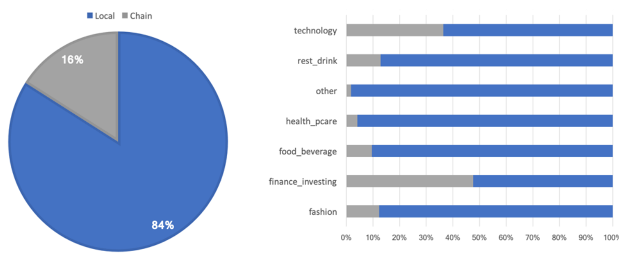


Figure 5: The ratio of local vs chain stores in general and for each class in the ShoS dataset

has a higher ratio of chain stores compared to the remaining classes and that is due to the dominance of large banks in the collected data.

Finally, the "difficulty" attribute was based on a pre-defined scales for some environmental and confounding factors. To explain, image-related issues, such as bad angle, unclear/blurred view, and poor resolution were assigned the value=1. The second difficulty scale, with the value=2, was assigned to store names that do not indicate any semantic clue regarding its type. Another difficulty scale defined with the value=3 was assigned to stores with misleading names. Finally, storefronts that are crowded with advertisements and non-signage elements were assigned the difficulty value=4. About 16% of the ShoS signboard images are difficult which increases the challenge of the ShoS dataset.

The ShoS dataset was revised and cleaned following specific protocols. First, the hired freelancers were provided with a guideline to follow during data annotation. For example, all text that appeared on signboards should be recorded manually except numbers and addresses as they do not add valuable information. The "difficulty" and "local" attributes were annotated by the researchers at the end to ensure uniformity. After that, the annotation file was processed using a Python script to avoid null values and mismatched attributes. Finally, the whole annotation file was revised manually over multiple stages to validate store class and sign text attributes.

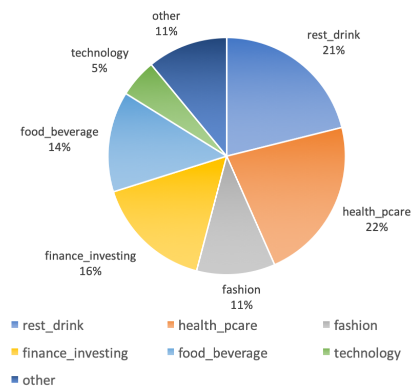


Figure 6: Distribution of store classes in the ShoS dataset

4. EXPERIMENTS

4.1 Shop Classifier

Based on the literature, NLP machine learning techniques can be used to classify text into some predefined classes. In this work, the traditional methods were preferred to deep learning because of the reasons mentioned earlier. After comparing a couple of methods, we decided to use Multinomial Naive Bayes MNB and Support Vector Machine SVM as they stand out among the other reviewed methods. According to our previous work the classifier was built and trained as follows [6].

The extended attributes recorded for the ShoS dataset were used for classification. In particular, the textual information recorded for each signboard (i.e. attribute: sign text) was utilized where each instance is associated with one of the six predefined store classes (i.e. attribute: class). The distribution of all store classes in the ShoS dataset is illustrated in FIGURE 6. To classify stores based on the textual information, several phases were implemented: text cleaning, feature extraction, and training/testing the models. FIGURE 7 shows the phases of store classification where the attributes “sign text” and “class” from the ShoS dataset were fed to classify shops. In the following discussions, each signboard text sample is called a document.

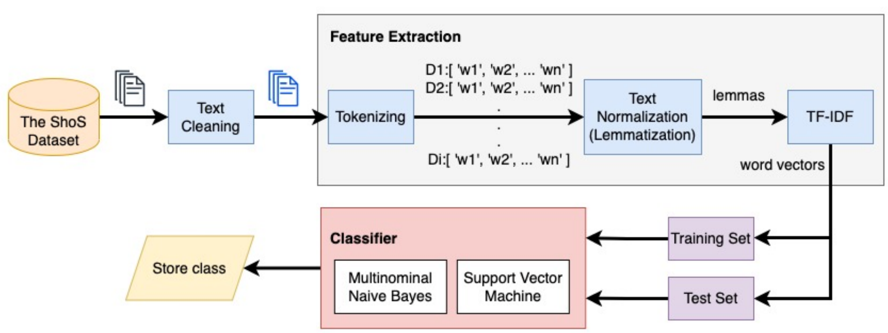


Figure 7: The phases of store classification

Ground Truth Text	After Cleaning & Processed Text
New Look LASER MEDICAL	new look laser medical
CIRCLE SUSHI & GRILL Dine in Carry out Delivery	circle sushi grill dine carry delivery
ALDO	aldo
Bench. FACTORY STORE	bench factory store
carter's	carter
Bank of America	bank america
SPACES.	space
The Tile Shop	tile shop
veruca chocolates	veruca chocolate
West Marine	west marine

Figure 8: A comparison of text data prior to and after cleaning and feature extraction

Since documents can contain some noise, the following cleaning steps were performed to enhance the quality of the text data. First, we converted the text data to lowercase. Then, we removed superfluous text data like punctuation, stop-words, and words with two or fewer characters using NLTK library. Next, we removed numbers from the text data as they are insignificant for the classification process. No spelling correction was needed as text data in signboards does not necessarily follow correct spelling. Finally, all duplicate documents, which is possible because of the existence of chain stores and other common naming, were removed. Therefore, the document size was reduced to 8904 after cleaning.

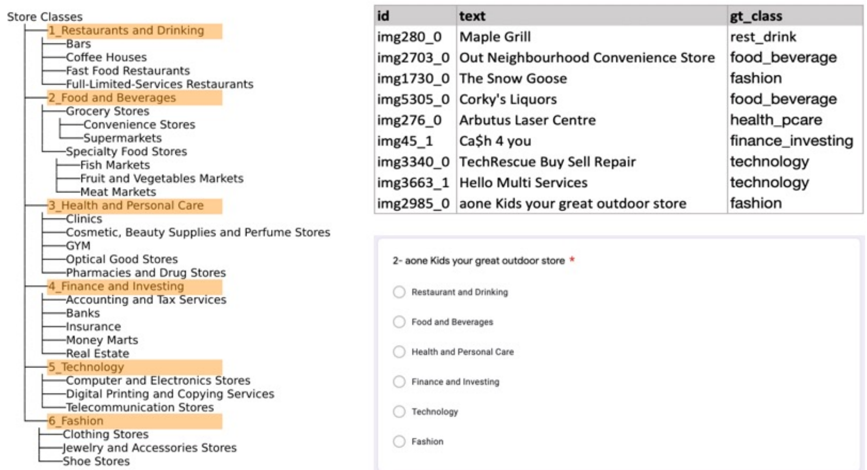
The feature extraction process started with tokenizing all words in each document in the ShoS dataset. Then, a text normalization technique was applied to prepare the text data for the classification process. Stemming and Lemmatization approaches are usually used with NLP problems in order to reduce morphological variations of words. The difference between them is that stemming removes the last few characters of a word and usually results in incorrect meaning and spelling, which is called Stem. On the other hand, Lemmatization considers context by converting the word to its meaningful base form, which is called lemma. In this research, the Lemmatization technique was applied as its accuracy is paramount and the ShoS dataset is not huge compared to NLP problems. FIGURE 8 shows a comparison of text data prior to and after cleaning and pre-processing.

Since each document in the corpus of the ShoS dataset belongs to one class only, we wanted to quantify words and assign a weight to each word in order to focus on significant keywords that carry a value for each store class. To implement that, the Term Frequency-Inverse Document Frequency TF-IDF technique was used. TF-IDF vectorizes all the text data at a word-level in order to classify the documents. The word vector is computed using the following equations 1, 2, 3, 4, where t is the word, d is the document (set of words per signboard), N is the count of corpus, and $corpus$ is the total document set.

$$TF\text{-}IDF(t,d) = TF(t,d) \times IDF(t) \tag{1}$$

$$TF(t,d) = \frac{count\ of\ t\ in\ d}{number\ of\ words\ in\ d} \tag{2}$$

$$DF(t) = occurrence\ of\ t\ in\ documents \tag{3}$$



id	text	gt_class
img280_0	Maple Grill	rest_drink
img2703_0	Out Neighbourhood Convenience Store	food_beverage
img1730_0	The Snow Goose	fashion
img5305_0	Corky's Liquors	food_beverage
img276_0	Arbutus Laser Centre	health_pcare
img45_1	Ca$h 4 you	finance_investing
img3340_0	TechRescue Buy Sell Repair	technology
img3663_1	Hello Multi Services	technology
img2985_0	aone Kids your great outdoor store	fashion

Figure 9: The class tree provided to the participants in the human comparison (left), and Some of the test set documents along with a sample from the survey (right)

$$IDF(t) = log(\frac{N}{DF+1}) \tag{4}$$

The vectorized data was split into training and testing sets applying 70/30 ratio. The training set was fed into the classifiers MNB and SVM. The “Other” class was excluded to avoid confusing the classifiers as samples of that class do not share common features. For the hyper parameters for MNB alpha=1.0, class_prior=None, and fit_prior=True; while in SVM, the weight of all classes was set by default to one, RBF kernel was used, squared hinge for loss function with 1000 max iteration and a penalty equal to l2 with no random state. The final result is a store class, and it is considered correct if it matched the ground truth.

4.2 Human Survey

The classification performance was compared with human performance using an online survey. An equal number of signboards from each store class were randomly selected from the test set of the ShoS dataset. The samples included 50% difficult ones with non-descriptive text for each class. This insured that the survey had a similar level of difficulty to our classification experiment. An online survey, built using Google Forms[3], was set to collect human responses on the samples where each participant had to classify text samples based on the class tree they were supplied with. FIGURE 9 shows the class tree provided to the participants, and some of the test set documents along with a sample from the survey. At the beginning of the survey, the participants were provided with the purpose of the study in addition to simple instructions regarding how to classify the text. If the participant was not able to determine the designated class, he/she was guided to choose based on their best guess. The survey took about 10 minutes to complete.

[3] https://www.google.ca/forms/about/

Furthermore, the quality of the responses was assessed based on two factors derived from the collected personal information. The first factor was the participants' level of proficiency in the English language. All responses related to participants with an English proficiency level of beginner or lower were excluded. The second factor was the length of the participants' living experience in Canada and the US. All responses of participants who lived in Canada and the US less than 6 months were also excluded. This way we avoided any invalid assessments by eliminating the outliers. The survey was distributed online through various communication applications.

5. RESULTS AND DISCUSSION

The results of testing MNB and SVM classifiers to classify shops based on the text appeared on their signboards reached an accuracy of 85.74% and 90.01% respectively. TABLE 2 shows the precision, recall, and f1-score for all classes for both classifiers. Since F1-score is the most preferable measure for such problems, we illustrated in FIGURE 10 a comparative bar chart for each class in both classifiers.

Table 2: The results of the store classification stage for MNB and SVM classifiers for all the studied classes

Class	**Evaluation metrics**	**MNB**	**SVM**
rest_drink	Precision	91%	80%
	Recall	88%	94%
	F1-score	90%	87%
food_beverage	Precision	92%	92%
	Recall	88%	90%
	F1-score	90%	91%
health_pcare	Precision	72%	96%
	Recall	97%	93%
	F1-score	83%	94%
finance_investing	Precision	92%	97%
	Recall	96%	96%
	F1-score	94%	96%
fashion	Precision	97%	87%
	Recall	50%	68%
	F1-score	66%	76%
technology	Precision	99%	95%
	Recall	67%	89%
	F1-score	80%	92%

It is noticed that SVM works better even with classes that have fewer samples and more non-descriptive text like class "Fashion" and "Technology". By looking at F1-scores of the SVM, it is observed that class "Fashion" has the lowest performance and that might be because of the high possibility of non-descriptive names used in their signboards as it represents 30% of the non-descriptive samples. In contrast, the class "Finance and Investing" has a stronger performance because of the consistency and limitations of the vocabulary that could be used on their signboards.

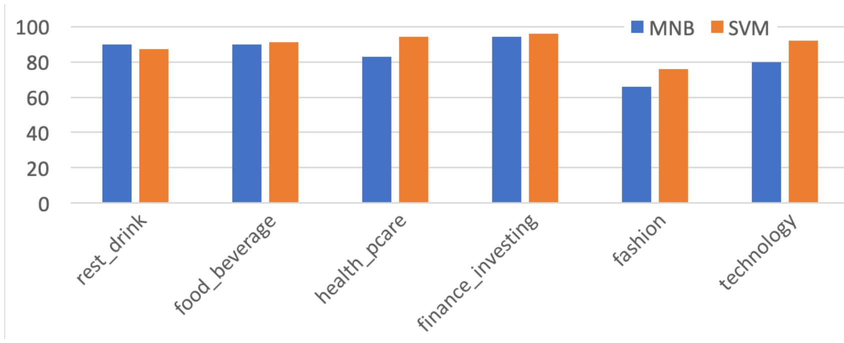


Figure 10: The F1-scores for MNB and SVM classifiers

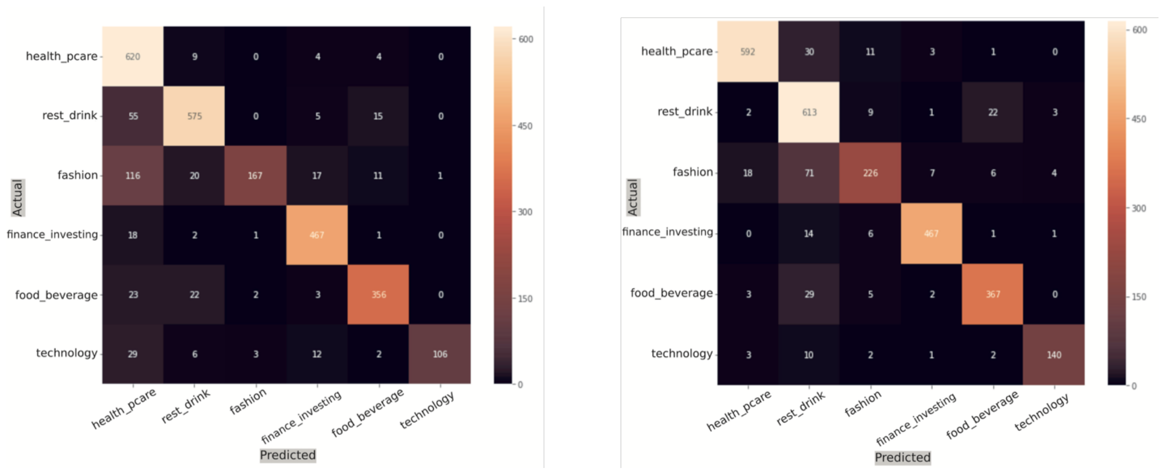


Figure 11: The Confusion matrix heatmap for MNB (left) and SVM (right) classifiers

FIGURE 11 illustrates the confusion matrices for both classifiers. The macro F1-scores for both classifiers are 83.83% and 89.33% for MNB and SVM respectively. When manual verification was performed on some samples of the confusing cases, it was observed that the confusion was caused by the usage of misleading words that are unrelated to store class such as “Nail Bar”, or because of the common vocabularies between two classes like the word “food” in “rest_drink” and “food_beverage” classes.

For the human survey, a total of 101 responses were collected. Females represented 52.4% and males represented 47.6% of the participant population. The survey results were analysed based on three measures: precision, recall, and F1-score. The results for all classes are included in TABLE 3 where the same samples exposed to the human were used to test our classifier. An illustration of the results for each store class is illustrated in FIGURE 12. Our classifier outperformed human performance by about 15% where it reached an F1-score=87.9% compared to 71.85% for human. It was observed that the ambiguity factor for non-descriptive text resulted in the majority of the misclassified text by human participants. This highlights the importance of adding descriptive keywords related to store class in the signboards as it will improve the ability of both humans and machines to classify stores accurately especially when the store façade is not representative. Moreover, the only class human was able to achieve slightly better score than our classifier was “Restaurant and Drinking” and that

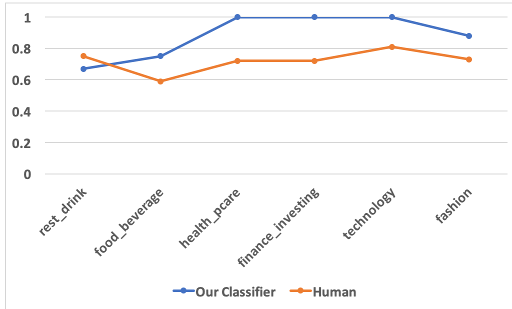


Figure 12: F1-scores for each store class for our classifier and human

was because of the common vocabularies between the mentioned class and "Food and Beverages" class. These similarities confused the classifier in most of the cases while human was able to get the semantic and differentiate them.

The results of the experiments and human survey highlighted the importance of adding descriptive and discriminative keywords related to store class in the signboards. That increases the accuracy of classifying stores by humans and machines especially when the store façade is not representative. This study is limited to the six shop classes, and that provides potential to extend the study to other classes.

Table 3: The overall performance measures to compare our classifier versus human for all store classes

	Precision	**Recall**	**F1-Score**
Human	71.92%	71.97%	71.85%
Our Classifier	**89.17%**	**87.5%**	**87.9%**

6. CONCLUSION

In this work, we extended the annotation of the ShoS dataset to include more attributes for shop classification purposes. Then, two models were trained and tested utilizing the ShoS dataset. SVM showed great performance even with classes that have a lower number of samples and a high number of non-descriptive text. The classification performance was compared with human performance, and the results showed that our classifier excelled over human performance by about 15%. Most of the misclassified samples were connected with the ambiguity factor. This highlights the importance of adding descriptive keywords in the signboards to increase the accuracy. For future work, more analysis can be done to enhance the classification performance, such as using augmentation techniques. Also, more store classes may be added for models' generalization.

7. ACKNOWLEDGEMENT

This research was supported by SEU (Saudi Electronic University) and NSERC (Natural Sciences and Engineering Research Council of Canada).

References

[1] Vivian Cook. The Language of the English Street Sign. Multilingual Matters, 2022.

[2] https://www.westmount.org/wp-content/uploads/2014/07/7-Storefronts_and_Signage.pdf

[3] Almuhajri M, Suen C. Intensive Survey About Road Traffic Signs Preprocessing, Detection and Recognition. In International Conference on Computing, Riyadh, Saudi Arabia, Springer. 2019:275-289.

[4] Al-qudah R, Suen CY. Impact of Font on Computer Recognition of License Plates On Automobiles. In Proceedings of the 3rd International Conference on Vision, Image and Signal Processing, ICVISP 2019, New York, NY, USA, 2019. Association for Computing Machinery.

[5] Almuhajri M, Suen CY. Shop Signboards Detection Using the ShoS Dataset. In El-Yacoubi MA, Granger E, Yuen PC, Pal U, Vincent N, editors, Pattern Recognition and Artificial Intelligence - Third International Conference, ICPRAI 2022, Paris, France, June 1-3, 2022, Proceedings, Part II, volume 13364 of Lecture Notes in Computer Science, Springer, 2022:235-245.

[6] Almuhajri M, Suen CY. A Complete Framework for Shop Signboards Detection and Classification. In International Conference on Pattern Recognition ICPR 2022, Montreal, Canada, August 21-25, 2022 (in-press). IEEE, 2022.

[7] https://arxiv.org/abs/1512.05430

[8] Movshovitz-Attias Y, Yu Q, Stumpe MC, Shet V, Arnoud S, Yatziv L. Ontological Supervision for Fine Grained Classification of Street View Storefronts. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA. 2015:1693-1702.

[9] Noorian SS, Qiu S, Psyllidis A, Bozzon A, Houben GJ. Detecting, Classifying, and Mapping Retail Storefronts Using Street-Level Imagery. In International Conference on Multimedia Retrieval (ICMR), Dublin, Ireland, 2020:495–501.

[10] https://arxiv.org/abs/1804.02767

[11] Karatzas D, Shafait F, Uchida S, Iwamura M, Bigorda LG, et al. ICDAR 2013 Robust Reading Competition ICDAR. In 2013 12th International Conference on Document Analysis and Recognition, Washington, DC, USA, IEEE. 2013:1484-1493.

[12] Karatzas D, Gomez-Bigorda L, Nicolaou A, Ghosh S, Bagdanov A, et al. ICDAR 2015 Competition on Robust Reading. In 2015 13th International Conference on Document Analysis and Recognition (ICDAR), Nancy, France. IEEE. 2015:1156-1160.

[13] https://arxiv.org/abs/1601.07140

[14] Yuan TL, Zhu Z, Xu K, Li CJ, Mu TJ, et al. A Large Chinese Text Dataset in the Wild. Journal of Computer Science and Technology. 2019;34:509-521.

[15] Wang K, Belongie S. Word Spotting in the Wild. In European Conference on Computer Vision, , Crete, Greece. Springer. 2010:591–604

[16] https://arxiv.org/abs/1903.10412

[17] Kadhim AI. Survey on Supervised Machine Learning Techniques for Automatic Text Classification. Artificial Intelligence Review. 2019;52:273-292.

[18] Kowsari K, Meimandi KJ, Heidarysafa M, Mendu S, Barnes L, et al. Text Classification Algorithms: A Survey. Information. 2019;10:150.

[19] Fanny F, Muliono Y, Tanzil F. A Comparison of Text Classification Methods K-NN, Naïve Bayes, and Support Vector Machine for News Classification. Jurnal Informatika: Jurnal Pengembangan IT. 2018;3:157-160.

[20] Rahman Mat, Akter YA. Topic Classification From Text Using Decision Tree, K-NN and Multinomial Naïve Bayes. In 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), IEEE, 2019:1-4.

[21] Octaviani NL, Rachmawanto EH, Sari CA, Rosal IMSD. Comparison of Multinomial Naïve Bayes Classifier, Support Vector Machine, and Recurrent Neural Network to Classify Email Spams. In 2020 International Seminar on Application for Technology of Information and Communication (iSemantic), IEEE. 2020:17-21.

[22] Dutta A, Zisserman A. The via Annotation Software for Images, Audio and Video. In Proceedings of the 27th ACM International Conference on Multimedia, MM ’19, New York, NY, USA, 2019. ACM.