Automated Triaging Medical Referral for Otorhinolaryngology Using Data Mining and Machine Learning Techniques

CHEE KEONG WEE[1,2], **XUJUAN ZHOU**[1], **(Member, IEEE), RAJ GURURAJAN**[1],
XIAOHUI TAO[3], **(Senior Member, IEEE), JENNIFER CHEN**[4], **RASHMI GURURAJAN**[4],
NATHAN WEE[5], **AND PRABAL DATTA BARUA**[1,6,7]

[1]School of Business, University of Southern Queensland, Toowoomba, QLD 4350, Australia
[2]Digital Application Services, eHealth Queensland, Brisbane, QLD 4001, Australia
[3]School of Mathematics, Physics and Computing, University of Southern Queensland, Toowoomba, QLD 4350, Australia
[4]Royal Brisbane and Women's Hospital, Queensland Health, Herston, QLD 4029, Australia
[5]Dialog Pty Ltd., Brisbane, QLD 4101, Australia
[6]Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW 2007, Australia
[7]Cogninet Brain Team, Sydney, NSW 2010, Australia

Corresponding authors: Chee Keong Wee (ckwee@outlook.com) and Xujuan Zhou (xujuan.zhou@usq.edu.au)

ABSTRACT Public hospitals receive and triage a large volume of medical referrals for otorhinolaryngology annually and it can be a challenge to derive knowledge from them as they are written in unstructured text and may be unavailable in electronic formats. Acquiring knowledge and insights from these referrals are important to public health management and policymakers. Triaging of general practitioner (GP) referrals for ear, nose, and throat (ENT) specialists is a manual process performed by experienced clinicians, but it is time-consuming. This paper proposes utilising machine learning and data mining to automate the process of referrals. In this study, an ensemble of machine learning algorithms to perform clinical text mining against the unstructured referral text in order to derive the relationship among the discovered medical terms was proposed and implemented. A set of comprehensive term sets' association rules which describe the entire referral dataset's characteristics was obtained from the association rule mining experiments. The neural network-based text classification model that can classify referrals with high accuracy was developed, tested and reported in this paper.

INDEX TERMS Artificial intelligence application, association rules mining, healthcare application of AI, machine learning and text classification, medical natural language processing, neural network.

I. INTRODUCTION

Patients are normally referred to specialists when the general practitioner (GP) requires further diagnostic support or is not able to provide specialised care for them. Another purpose of the referral process is to support the GP in managing the care that the patients have received and keep track of their progress [1], [2]. Every year, there are millions of medical referrals written by general practitioners to refer their patients to other medical professionals for specialised treatment in the state of Queensland [2]. Healthcare organisations such as hospitals and medical centres triage patients' medical conditions based on the referrals to ensure that they can receive

the relevant and timely care to avoid care duplication and untreated illness, thus minimising cost and resources [2], [3]. These referrals are designed to help patients to obtain the appropriate level of care. They contain the current conditions, including the patient's medical history and list of medications [4].

However, not all specialists are within the GP's consideration when they decide to refer the patients. Each GPs has their referral circle, specialists that they are familiar with, or have worked with in the past. This can create an element of bias, in that the referrer may ascribe a particular diagnosis to a set of symptoms or may assign a particular level of urgency to the referral due to patient pressures. The objective of triaging is to get the right patient to the appropriate specialist for the first time. The process of triaging referrals is performed

The associate editor coordinating the review of this manuscript and approving it for publication was Mu-Yen Chen.

by certified clinicians (usually a nurse or doctor) to sort and allocate the patients following a system of priorities to specialists to shoehorn the quality of care that they need and their survival rate, whilst minimising delays and cost [5]. The triaging clinicians decide which referrals should be prioritised, by examining the level of urgency and the pathway for the patients [6], [7]. Cases that are deemed to be clinically serious or severe will receive higher priorities on the triage list and will therefore see the specialist more urgently. Most health organisations that employ the triaging process categorise the referrals into Category 1 (urgent), 2 (moderate urgency) and 3 (non-urgent). Medical organisations employ a large group of clinicians and systems to manage the triaging process.

The majority of the referrals are sent in electronic documents such as word or Portable Document Format (PDF) files and they are written in free text by the GPs describing the patient's condition or challenges in managing the presentation [2]. The current method of processing general practitioner (GP) referrals has been computerised with referrals coming in digital formats with PDF and scanned images as examples. However, the triaging procedure is still performed by clinicians, in which each document is reviewed manually [1]. Every GP has their personal style of writing referrals, and there is wide variation in the content and structure in the referral formats [3]. Each clinician who triages referrals also has varying levels of professional experiences and familiarity with the clinical operation [5]. While there are governing policies and procedures for triaging, it is not unusual for variations among the clinicians' approach to triaging and associated biases [3], [6], [7].

Beside triaging, healthcare management parties also need to gain clinical insights into the large volume of medical referrals that are received yearly in all medical specialties [3]. Examples such as identification of medical entities described among the referrals and under different triaged groups, their associative relationships that constitute the referred contents for the patients, and their frequency of occurrence are some of the knowledge that can be acquired [6]. This allows the healthcare management and policymakers to appreciate the variation among the referrals dataset's contents as well as the clinicians' teams triaging expertise [2]. Several potential benefits can be gained from achieving this; namely variance can be reduced and biases can be adjusted for in calibrating the criteria and policy that govern the triaging process and gauge the triaging clinicians' judgments as a collective [3].

While conventional Natural Language Processing (NLP) can mine contents from unstructured text, they are usually not equipped to handle medical content. Therefore, new branches of medical-based NLP have been repurposed in recent times to handle the Unified Medical Language System (UMLS) [7]. The medical referrals contain medical terms which details information about the patient's condition, history, illness, location, interventions already performed, and medications prescribed. These terms can be represented as codes from the International Classification of Diseases (ICD) [5]. The diagnoses are written in unstructured text formats [6] which makes it challenging to extract useful content. This feature engineering can also open other avenues to pursue clinical insights such as outlier detection among the referrals [8]. Furthermore, conventional text classification is unable to score effectively against ICD due to the characteristics of the medical text contents.

The process of triaging involves several stages: firstly, data extraction of the referrals, which arrive in numerous formats from binary files like pdfs to physical copies via fax or mail. Secondly, review of the content to ascertain their validity, domain, and relevancy in the referrals. The triaged referrals are then assessed based on a set of rules to determine their urgency [7]. Finally, the allocation of pathways or services. The triaged referrals are then assigned to the pathway following the clinicians' judgment. The model of the triage process is reviewed manually by the clinicians periodically to ensure that they are up to date and reflect contemporary practices [6], [7].

The goal of this research is to support public health policymakers to be informed and have better insight into the triaged referrals' landscape so that policies can be improved through effective quality control over the triaging processes. The specific objective of this research is multifaceted: firstly, to develop a novel approach to triage GP's referrals. Secondly, to acquire knowledge and insights about the medical entities and their associative relationships that make up their occurrence among the various categories of the triaged referrals from the large volume of referrals and to automatically classify them. Thirdly, to apply the concept of outlier detection to identify referrals that are considered as outliers in each triaged group.

To achieve the goals, this paper proposes a novel approach of mining and classifying the large pool of GP referrals and acquiring insights into the contents using a combination of a medical-capable NLP service [1], association rule mining [3] and a combination of dimensionality reduction with outlier detection [8], [9]. For the dataset, we obtained a group of GP referrals that are from the medical speciality of Otorhinolaryngology. One of the main medical NLP used for this research is AWS (Amazon Web Services) Comprehend Medical [10] which we extract and encode the notes' entities into proper terminology [4].

This research makes several significant contributions:

1. The application of machine learning/deep learning to real world applications (medical referrals) in order to acquire a high level of accuracy with evidence.
2. Support of the Queensland public health system in the potential application of machine learning/deep learning to provide decision support to clinicians.
3. Establishment of a novel approach that enables clinicians and policymakers to appreciate the landscape of medical needs and deficits by creating useful clinical insights with the use of data mining and machine learning.

The rest of the paper is organized as followed: Section II briefly presents a summary of related works including technologies and algorithms. III describes the overview of the three designs, covering the proposed approach. Section VI discusses the study dataset. The empirical results and analysis are reported in section V. The last section provides a conclusion of this study with suggestions on future enhancements and potentials.

II. RELATED WORKS

To achieve the goal of this research, three AI-based models are proposed; Referral Classification, Referral Insight and Outlying Referral detection. In this section, we first give a brief description of the data mining, machine learning/deep learning techniques, outlier detection and medical NLP techniques used for building these three AI-based models and then summarise some recent medical referrals works.

A. ASSOCIATION RULES MINING

Association Rules Mining (ARM) is a data mining procedure that discovers the relationship among the items in a dataset that are based on rules that can have some measure of interest. It is commonly known as Market Basket Analysis. According to Agrawal *et al.* [14], the rule to measure the interest between two items, X and Y, from a dataset, N, can be expressed as *Support*, *Confidence* and *Lift*. For the rule of $X \rightarrow Y$, the *Support* indicates the frequency of the itemset in the dataset, the *Confidence* is the frequency of the rule that is found to be accurate, and the *Lift* of the rule represents the ratio of the rule's support to the support of X and Y when they are independent.

$$support(X \rightarrow Y) = \frac{frequency(X, Y)}{N}$$
$$confidence(X \rightarrow Y) = \frac{frequency(X, Y)}{frequency(X)}$$
$$lift(X \rightarrow Y) = \frac{support(X \cup Y)}{support(X).support(Y)}$$

The Apriori algorithm was proposed by Agrawal and it has a bottom-up approach in identifying the items that are the subset within the dataset that meet the given threshold of C transactions [14]. The algorithm then extends the frequent incrementally, also known as candidate generation, and then repeat the process in testing the groups of candidates against the dataset to discover more rules.

B. DEEP LEARNING METHOD

A various of machine learning/deep learning algorithms have been applied to medical and healthcare domain including cancer research [37], prediction of sepsis [38], chronic pain detection [39], [40], coronary artery disease research [41]–[43].

A neural network is a group of machine learning algorithms that mimic the organic human brain to recognize a set of data's underlying relationship through the process of minimizing a loss function via experiencing with the learning of weights on the neurons [11], [12]. The neural network consists of layers of interconnected nodes known as perceptron and each perception receives multiple signals produced from data into an activation function which produce output that is fed into other perceptron or as result. Deep learning is another implementation of a neural network where multiple layers of perceptrons are arranged in an interconnected layout where the input layer received data input patterns while the output layer produces corresponding classification or output signals. The layers between the input and the output layers are referred to as hidden layers and it optimizes the weighting of the input within the layers until the neural network's error margin is minimal [11], [12].

C. UNIFORM MANIFOLD APPROXIMATION AND PROJECTION (UMAP) AND LOCAL OUTLIER FACTOR (LOF)

Uniform Manifold Approximation (UMAP) is a dimensionality reduction algorithm that is based on manifold learning techniques such as Riemannian geometry, and topological data analysis [8]. The GP referral dataset is represented as a matrix and it is considered as high dimensional data which that is difficult to process [8], [13]. Such high dimension dataset must be reduced to mitigate the curse of dimensionality, to a lower dimension for easier analysis. UMAP's approach is to present the data on an embedded non-linear manifold on a space of higher dimension, then the data is visualized in space of lower-dimension once the manifold is reduced to a lower dimension [8], [13]. it is a neighbour-graph algorithm that defines fuzzy simplicial sets using local metric spaces when are concatenated to form a large singular topological structure. A k-value is provided by the user and it is used to balance the local-global structure representation, which is comprised of the local metric spaces that are calculated from each data points to their k-th nearest neighbour [8]. This is followed by the embedding process which is a low dimension layout that is built from the fuzzy set cross-entropy that matches the largest edge weights from the structure's topology using the minimum pairwise distances available in the layout, and this process is reversible [8], [13].

Once the data's dimensionality is reduced, the next step is to perform the outlier detection. For this, the Local Outlier Factor (LOF) algorithm is used [9], [14]. The LOF algorithm computes the local density deviation for selected data points against their neighbouring data points and regards those data points that have a much lower density value to be outliers. It iterates through all the data points and finds its k-distance from its neighbouring data points in Manhattan distance which is then used to derive its Local reachability density (LRD) [9], [14]. The value of the surrounding datapoint's LRD is used to determine the Local outlier factor (LOF). And whichever datapoint has the highest LOF value is then considered as an outlier [9], [14].

D. MEDICAL NATURE LANGUAGE PROCESSING

Medical NLP have acquired significant improvement by many researchers and especially cloud-based service providers to decipher large volume of unstructured medical texts [15]. Sources for these medical text are ubiquitous among medical centres and hospitals and the medical NLP acquire not only the specific information but also to link them to other medical ontologies such as ICD, UMLS to serve other medical services such as clinical insights, finance, insurance, or serve as feature engineering to support other ensembled machine learning models for specific clinical application [17]. One of the common uses of Medical NLP is in the triaging of patients' conditions in places such as emergency departments, specialist clinics [1], [18]. Other used are creation of medical intelligence that enable clinicians to understand the patients' groups of conditions and status of their health [19], [20].

E. SUMMARY OF RECENT MEDICAL REFERRAL WORKS

Researchers have applied machine learning algorithms to a range of the medical disciplines in the healthcare sector [21]–[23].

A summary of recent studies that use machine learning models to triage patients across for medical services among the hospitals are displayed in Table 1.

The studies [24]–[27] are closely matched to the goal of this paper's objective where the urgency to attend to the patients' needs at the emergency departments by triaging their current medical conditions and history. In particular, the study [28] used convolutional neural network (CNN) and artificial neural network (ANN) to assist with the triaging of ophthalmology referrals. CNN provided the best accuracy of 81% on the test set and ANN achieved accuracy of 77%. Studies [29], [30] have similar objectives but they are targeting medical specifics which have similar triaging requirements. The studies use both statistical algorithms and deep learning to evaluate their effectiveness and their outcomes of their classifications are calculated on their precision, recall and specificity results. Their outcomes show that deep learning fair better than other models but require additional support in model training and tuning.

III. DESIGN OF THREE AI-BASED MODELS

Three AI-based models namely Referral Classification, Referral Insight and Outlying Referral detection are proposed to achieve the research goals. In this section, we described the design of these three models.

A. AI-BASED REFERRAL INSIGHT MODEL DESIGN

For this paper, the terms refer to the individual medical entities or terms that are identified among the referrals by the Medical NLP [1]. A term set refers to a series of medical terms present in a patient's referral and they have a significant associative relationship with each other. The proposed novel approach is called Referral AI Insight (RAII) and it is shown in Fig. 1.

It uses a combination of medical capable NLP and Association Rules algorithms [13] to mine the GP referrals' contents within 4 distinct stages of operation. Due to the sensitivity of the data, some part of the referral's contenting parsing is done by a separate group of information technology (IT) specialists to ensure data security and privacy.

The first stage – All the GP referrals are in PDF format and are uploaded to the cloud blob storage. Their contents are extracted out via Python's API (Application Programming Interface) libraries and stored in a cloud SQL database. Each record has a triaged label that the clinicians have assigned to them. Second stage – The referral contents are then parsed by the Medical NLP service which identifies and retrieves out the medical entities, or terms, as well as their associated medical-ontology linked codes (ICD10CM and RxNorm). The medical NLP provides a list of possible matching codes together with their confidence score. We select only those codes that have the highest score and are above the acceptable threshold of 50% confidence. The codes are then combined into an array and stored in the same database. The third stage - Once the 2nd stage is complete, we retrieve all the records' parsed medical codes under a specific referral category and then apply count vectorization on it to create a large sparse matrix. In the fourth stage – the Apriori algorithm is applied to the dataset.

Results obtained are a comprehensive list of rules between the itemset of medical codes from the dataset, which are then plotted and discussed later in the paper.

B. AI-BASED REFERRAL CLASSIFICATION MODEL DESIGN

Our AI-based Referral classification model uses a combination of medical NLP and deep neural networks to classify triage GP referrals as shown in Fig. 2. These processes are comprised of 4 stages and they are performed by different groups of IT specialists to ensure data security and privacy. The first stage is to upload the GP referrals in PDF format onto the blob storage and extract their contents out via pdf API libraries. The second stage involves using the medical NLP services the extract referral contents and extracts out the medical entities as well as their respective medical codes. The third stage involves encoding and vectorizing the dataset's content medical codes and labels. Lastly, the final stage is to use the dataset and train the models.

The experiment includes using other machine learning models and compares the results with our main AI-based Referral Classification model. It is intended that our new solution becomes flexible enough to adapt to any medical domain with greater flexibility and accuracy as compared to other generic text classification methods for healthcare [4]. As the tests are conducted against a sample set of the GP referrals that have been triaged by experienced clinicians, the results are validated against their original classifications. To our best knowledge, these are the first methods that use both healthcare capable NLP services

TABLE 1. Summary of recent medical referral works.

| Authors, Year, Reference no.	Approach/Model	Aim	Dataset
Hong, *et.al.*, 2018, [23]	Neural network, gradient boosting (XGBoost), logistics regressions	Triaging admission of patients at the hospital's emergency department based on their existing conditions and medical history	972 Electronic medical records' Dataset from Yale New Haven Health system.
Raita, *et.al.*, 2019, [24]	Lasso regression, random forest, dee neural network, gradient boosted decision tree	Emergency department triage prediction of clinical outcomes	Ambulatory health care data (2007-2015) of age >18 patients, from National Centre for Disease Control and Prevention (CDC)
Goto, et.al., 2019, [25]	Lasso regression, random forest, gradient-boosted decision tree, and deep neural network	To predict clinical outcome for children during Emergency Department triaging using machine learning	52037 children medical records of ED visitation from National Hospital Ambulatory Medical Care Survey (2007 to 2015)
Horng, et.al., 2017, [26]	Logistic regression, naïve Bayes, and random forests	Trigger sepsis clinical decision support at emergency department using machine learning	230,936 patient visits to Emergency Department at Beth Israel Deaconess Medical Centre (BIDMC)
Tan, et.al., 2020, [27]	Convolutional neural network (CNN) and artificial neural network (ANN)	To assist with the triaging of ophthalmology referrals	208 participants from the Royal Adelaide Hospital (January 2018 to March 2019). 118 category one referrals, 61 category two referrals and 29 category three referrals.
Swaminathan, et.al., 2017 [28]	Gradient-Boosted Decision Trees, Logistics Regression	To triage patients that suffer from chronic obstructive pulmonary disease using machine learning	5600 OCPD patients records from six medical centres and medical universities in USA
Choi, et.al., 2019, [29]	Logistic regression, random forest, XGBoost	To triage patients at the emergency department for rapid treatment response using machine learning	142,972 patients record from Seoul National University hospital, south Korea
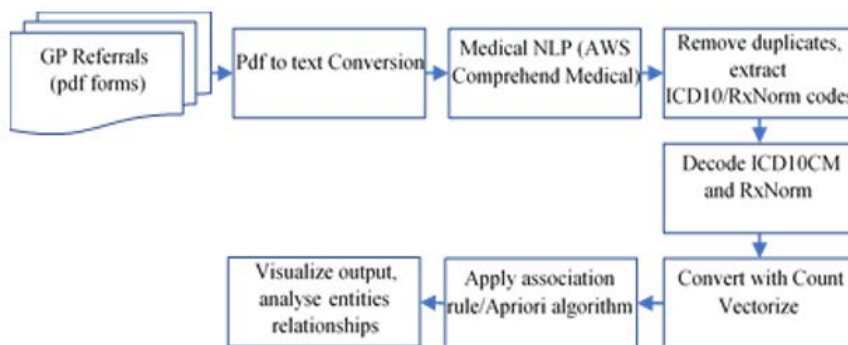
FIGURE 1. AI-based Referral Insight design–association rules mining.

that clinical terminology and deep learning for triaging GP referrals. The results from our experiment showed that this approach can yield high accuracy results in triaging medical referrals.

Referring to Fig. 2 and 3, the neural network model is built with 5 layers. Starting with the 1st layer with the equivalent number of neurons to match the input vector size, the second layer that follows is half the number of neurons of layer 1. The third layer's neuron number is arbitrary, and the fourth layer is set at 1/10 of the number. The last layer's neurons numbers match with the output (Y) size. The model is trained with 2395 samples and then validated on 267 samples. The activation functions for the neurons are set to Rectified Linear Function (Relu) and sigmoid to both the layer 1-4 and the final layer of neurons. The mode of classification is set to binary crossentropy and the optimizer to adam. We test with a combination of epoch and batch size for the model's training.

Based on the result as shown in Figure 3, the model starts to converge at the 10th epoch. The validation accuracy achieved in this setup is observed at 93% with the validation loss below 0.02.

C. AI-BASED REFERRAL OUTLIER DETECTION MODEL DESIGN

For our AI-based referral's outlier detection model that is illustrated in Fig. 4, both the UMAP and LOF algorithms are used to reduce the dimensionality of the referrals data and locate those data that are regarded as outliers. Using the same processed data from section C's approach, where the extracted medical terms are feature engineered by the AWS Comprehend Medical in ICD codes, they are also vectorized to produce a sparse matrix. This is in turn used by both UMAP and LOF to locate the outliers. The outliers' data comprise
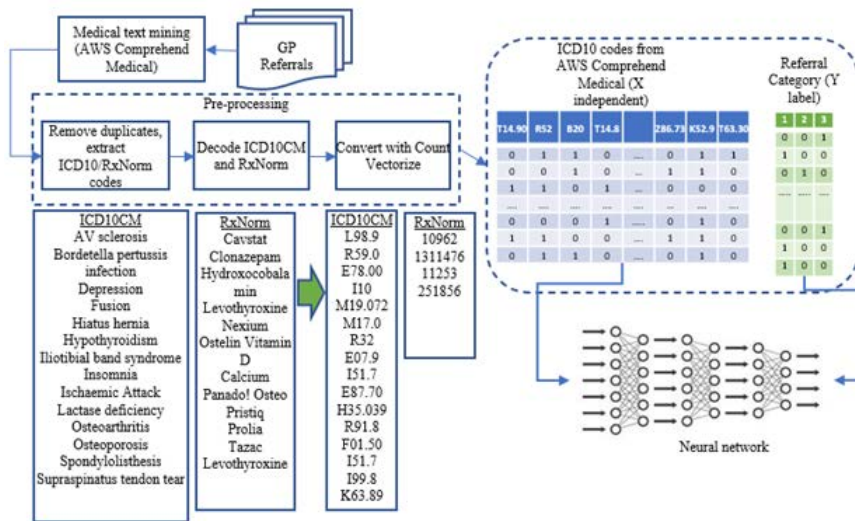
FIGURE 2. AI-based referral classification design.

```
Layer (type)                 Output Shape               Param #
=================================================================
dense_1 (Dense)              (None, 3206)               5142424
dense_2 (Dense)              (None, 1603)               5140821
dense_3 (Dense)              (None, 500)                802000
dense_4 (Dense)              (None, 50)                 25050
dense_5 (Dense)              (None, 3)                  153
=================================================================
```

FIGURE 3. Neural network's specification.

a list of ICD codes and we can reference them back to the original patients' referrals.

IV. STUDY DATASET

A. DATA SOURCES

In this experiment, the sample data is obtained from QLD Health's GP referral system for the field of otorhino-laryngology (ear, nose, and throat) with the date ranging from 2019 to 2020. The sampling of the data from the system complies with the States' safety and privacy of the patient's information which had been performed by respective clinical support staff. The dataset is then parsed by a separate IT application team using AWS Comprehend Medical [1] through their secured private virtual network with strong network encryptions. A subset of the samples is then made available to use in this experiment. The sample has a total of 3000 ENT referrals, and they have been labelled from the triaging process that has been conducted by the clinicians. The referral's label has a range from 1 to 3, with 1 signifying the highest urgency. There are 1000 records for each category to ensure an equal representation. The content of the referrals varies from different GPs, with some having more insights about the patients' present medical conditions, while others have included the patient's past medical history, and some have added in their current medications' routines. The machine in which this experiment is conducted is located

within the organization's secured IT computer system and protected by firewalls, security policies and network ring-fencing.

We use a cloud-based medical text processing facility that is specialized in the medical field called AWS comprehend Medical. It can derive useful information such as medical conditions, medical and protected health information (PHI) from medical notes using natural language processing [11], [12]. The detected entities are scored on the probability of their match through its processes of the ontology-connecting process to the standardized medical knowledge bases of RxNorm and ICD10-CM. Therefore, each referral is processed and decomposed to JSON text which contains the entities found and their closest matching RxNorm and ICD10-CM codes. We picked only those codes with the highest match score and tabulate them against the referral category.

B. DATA PRE-PROCESSING AND PREPARATION

The data pre-processing and preparation method is depicted in Fig. 5.

Normal human-readable texts are complex in structures, meaning and are constructed for any machine to interpret. Natural Language Processing (NLP) is a form of artificial intelligence that enables computers to read and interpret text data, thus giving them the ability to measure and
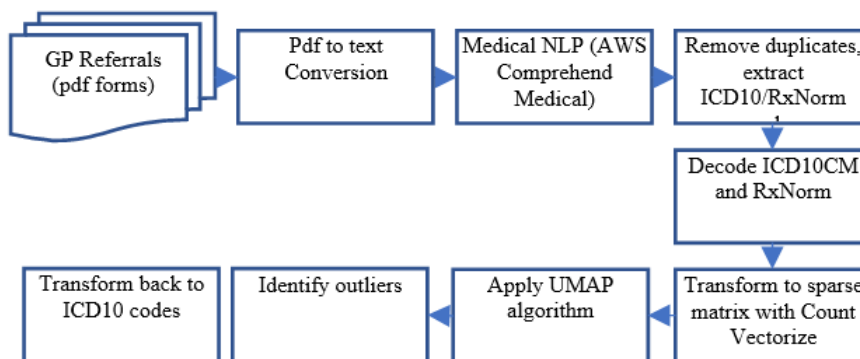
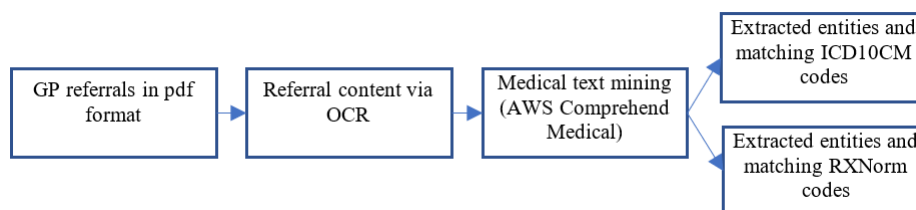**FIGURE 4.** AI-based referral insight design–outlier detection.

FIGURE 5. Data pre-processing and preparation.

determine the importance of specific text segments. The volume of Electronic Health Records (EHR) generated by healthcare-related IT systems posed a challenge for the clinicians to work with. NLP is gaining popularity in the healthcare sector as it can process these unstructured text data quickly to derive their meaning with great efficiency. But due to the nature of the business, there are numerous ways that certain important meanings may have ambiguities and inconsistencies. Examples include scenarios where ambiguities are usually caused by spelling errors, abbreviations, and description variations etc., where knowledge-base of normal language processing are not equipped to handle [15].

Text classification has been widely used to analyse and categorize texts across industry and academia [16]. Many models have been developed for this purpose such as linear regression, support vector machine, random forest, and deep learning to name a few. When it comes to performing text classification of medical text, there is a level of uncertainty especially when it comes to medical terminologies relating to illnesses, procedures, symptoms, drugs, prognosis, etc., where the lemmatization and stemming of conventional NLP knowledge bases are not capable to handle this [15]. This proved to be a challenge to perform text classification in the healthcare sector for a long time. However as of present, several cloud service and AI vendors have built specialized NLPs to cater to healthcare, and one example is AWS Comprehend Medical [17] which can derive the meaning from medical-related notes following HIPAA regulations. It can link the meanings of medical terms to the standard medical ontologies like ICD10-CM or RxNorm [1]. With

this feature, each medical note can be analysed by these medical NLPs to derive and match relevant entities to the closest medical ontology's codes. Consequently, each medical note can have an array of medical-related codes that are subsequently exploited by various forms of artificial intelligence.

In our Proof of Concept (POC) environment, the GP referrals are received in PDF file format and stored in blob storage. Their details are stored in a database with their medical contents processed and extracted by AWS Comprehend Medical, including their ontologies-link codes. It generates a list of suitable codes and their probability scores against each detected medical entity. So, our program will scan and pull only those code that has the highest scores with a confidence level above 60%. Fig. 5 illustrates the different stages in the data pre-processing. Tables 2 and 3 show a sample of the processed outputs from several referrals' contents. Table 4 shows a sample of the processed referrals where their contents of extracted ICD10 codes have been filtered out and merged into an array. The next stage is to prepare and encode the processed dataset before they are ready for the association rules mining and the neural network's training and testing programs. There are several types of encoding ranging from numerical representation to hashing and vectorization. The conventional vectorizing process will use tokenization to remove certain unimportant or less meaningful words, but the text data has been tokenized by AWS Comprehend Medical during the first phase. What the vectorization process contributes to this stage is to create a collection of all the medical codes from the data

TABLE 2. Processed sample based on ICD10CM concepts.

['Enlarged lymph nodes, unspecified (R59.9)', 'Disorder of the skin and subcutaneous tissue, unspecified (L98.9)', 'Other disorder of circulatory system (I99.8)', 'Localized enlarged lymph nodes (R59.0)', 'Disorder of brain, unspecified (G93.9)', 'Neoplasm of unspecified behavior of brain (D49.6)', 'Migraine, unspecified, not intractable, without status migrainosus (G43.909)']
['Localized swelling, mass and lump, right upper limb (R22.31)']
['Gastro-esophageal reflux disease without esophagitis (K21.9)', 'Diverticulum of esophagus, acquired (K22.5)', 'Visible peristalsis (R19.2)', 'Atrial premature depolarization (I49.1)', 'Unspecified abdominal hernia without obstruction or gangrene (K46.9)']
['Unspecified lump in the right breast, unspecified quadrant(N63.10)', 'Localized enlarged lymph nodes(R59.00)', 'Unspecified lesions of oral mucosa(K13.70)', 'Soft tissue disorder, unspecified(M79.9)', 'Abdominal tenderness, unspecified site(R10.819)', 'Cranial nerve disorder, unspecified(G52.9)', 'Localized swelling, mass and lump, neck(R22.1)', 'Other disturbances of skin sensation(R20.8)', 'Hemorrhage, not elsewhere classified(R58)', 'Pain, unspecified(R52)', 'Rash and other nonspecific skin eruption(R21)', 'Major depressive disorder, single episode, unspecified(F32.9)', 'Essential (primary) hypertension(I10)', 'Psoriasis, unspecified(L40.9)', 'Heart disease, unspecified(I51.9)', 'Old myocardial infarction(I25.2)', 'Chronic obstructive pulmonary disease, unspecified(J44.9)', 'Dorsalgia, unspecified(M54.9)', 'Zoster without complications(B02.9)', 'Disappearance and death of family member(Z63.4)', 'Family history of malignant neoplasm, unspecified(Z80.9)']
…………………

TABLE 3. Processed sample based on RxNorm concepts.

['Mupirocin 0.02 MG/MG Nasal Ointment [Bactroban] (108758)', 'Omega-3 Acid Ethyl Esters (USP) 1000 MG Oral Capsule [Lovaza] (607044)', 'Potassium Chloride 20 MEQ Extended Release Oral Tablet [K-Tab] (1483340)', '14 ACTUAT fluticasone furoate 0.1 MG/ACTUAT Dry Powder Inhaler [Arnuity] (1547664)', 'lisdexamfetamine dimesylate 30 MG Oral Capsule [Vyvanse] (854836)']
['candesartan cilexetil 16 MG Oral Tablet(577776)', 'Zenpep (1117097)', '12 HR Hyoscyamine Sulfate 0.375 MG Extended Release Oral Tablet [Symax]….(1046775)', 'osthol(1547630)', '24 HR Calcium Carbonate 200 MG / Folic Acid 1 MG / pyridoxine 75 MG / Vitamin B 12 0.012 MG Extended Release Oral Tablet(1087258)', '1 ML denosumab 60 MG/ML Prefilled Syringe [Prolia](993456)', 'Dextromethorphan Hydrobromide 2.5 MG/ML Oral Suspension(1536151)', 'Amiodarone hydrochloride 100 MG Oral Tablet(835956)', 'trimethoprim(10829)', 'aspirin(1191)', 'anti-inhibitor coagulant complex(314504)', 'Soma(204362)', 'Folic Acid 0.8 MG / Vitamin B 12 1 MG Disintegrating Oral Tablet(763597)', 'Methotrexate 2.5 MG/ML Oral Solution(1921592)']
['clopidogrel 75 MG Oral Tablet (309362)', 'Vitamin B 12 1 MG Oral Tablet (309593)', 'Cartia (491095)', 'Metoprolol Tartrate 50 MG Oral Tablet (866514)', 'Rosuvastatin calcium 20 MG Oral Tablet [Crestor]859753)', 'Allopurinol 100 MG Oral Tablet (197319)', 'Phymatolithon calcareum preparation (1731724)', 'crospovidone (19 MPA.S at 5%) (1661326)']
['tyrosine (10962)']

TABLE 4. Processed data with medical encoded entities and referred labels.

| Row	Referral Category	ICD10 list
0	1	I63_9, S02_2, Z18_12, W19, R04_0, M51_26, I44_…
1	1	T14_90, R52, B20, E11_9, T14_8, Z86_73, K52_9, …
2	1	L94_1
3	1	Z85_3, F32_9, J18_9, R74_0, E23_0
4	1	D72_1, T14_90, B99_9, B96_89, I25_2, C18_9, N7…
…	…	…
2657	3	J34_89, R06_89, J30_9, R09_81, K56_609, T78_40…
2658	3	Z86_19, J34_89, T85_698, J98_8, F50_9, Q95_0, …
2659	3	I51_9, G35, I65_29, H91_90, F79, F42_9, H91_93…
2660	3	G47_00, G47_30, G47_33, J22, M48_00, J40
266	3	G47_33, R52, I25_10, J00, R05, R09_3, F45_8, R…
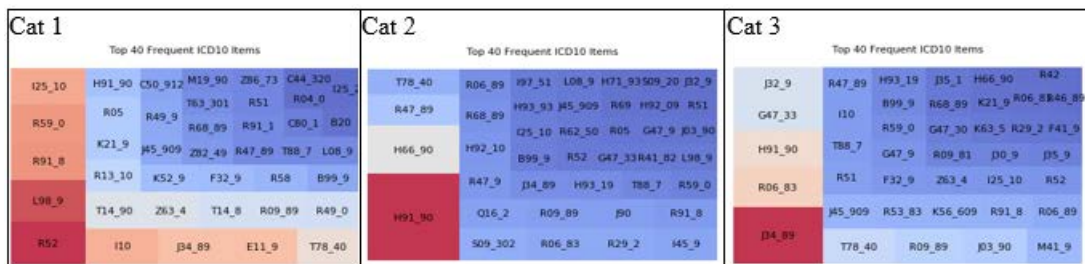
FIGURE 6. Most common medical entities/terms in all three referral categories.

that have been parsed by the AWS Comprehend Medical. Furthermore, it represents the text data via vectorization against the collection.

There are two methods of vectorizations: Count vectorization, where the word occurrence is represented by their counts, and Term Frequency–Inverse Document Frequency (TFIDF) vectorization, where the word occurrence is represented in the normalized form following the overall frequency of occurrence of text across the entire dataset.

For the deep learning model, the dataset is then split into a training and a testing group in the ratio of 80/20 and 90/10 for two sets of training and testing. The subset of the data is further segregated into input (X) and output (Y) sets followed

by the application of vectorization. The size of the X vector is 3206 x 2395 rows of input data. Consequently, the output data of the Y vector shares the same size.

V. THE RESULTS AND ANALYSIS
In this section, we report and analyze the empirical results from the AI-based referral insight model and AI-based referral classification model.

A. THE RESULTS OF AI-BASED REFERRAL INSIGHT MODEL FROM ASSOCIATION RULES MINING
The first set of results showed a heatmap of the top 40 most common medical terms' ICD10 codes in the three referral

TABLE 5. Samples of top 40 most common medical entities used in each category and their descriptions.

Category 1			Category 2			Category 3		
Entities	Item	Description	Entities	Item	Description	Entities	Item	Description
129	R52	Pain, unspecified	388	H91_90	Unspecified hearing loss, unspecified ear	269	J34_89	Other specified disorders of the nose and nasal sinuses
126	L98_9	Disorder of the skin and subcutaneous tissue, unspecified	216	H66_90	Otitis media, unspecified, unspecified ear	186	R06_83	Snoring
113	R91_8	Other nonspecific abnormal findings of lung field	139	R47_89	Other speech disturbances	175	H91_90	Unspecified hearing loss, unspecified ear
112	R59_0	Localized enlarged lymph nodes	111	T78_40	Allergy, unspecified	149	G47_33	Obstructive sleep apnea (adult) (pediatric)
108	I25_10	Atherosclerotic heart disease of native coronary artery without angina pectoris	108	S09_302	Unspecified injury of left middle and inner ear	149	J32_9	Chronic sinusitis, unspecified
106	I10	Essential (primary) hypertension	102	R06_83	Snoring	135	T78_40	Allergy, unspecified
101	J34_89	Other specified disorders of nose and nasal sinuses	101	R29_2	Abnormal reflex	125	R09_89	Other specified symptoms and signs involving the circulatory and respiratory systems
101	E11_9	Type 2 diabetes mellitus without complications	98	I45_9	Conduction disorder, unspecified	119	J03_90	Acute tonsillitis, unspecified
89	T78_40	Allergy, unspecified	95	Q16_2	Absence of eustachian tube	103	M41_9	Scoliosis, unspecified
82	T14_90	Injury, unspecified	91	R09_89	Other specified symptoms and signs involving the circulatory and respiratory systems	99	J45_909	Unspecified asthma, uncomplicated
82	Z63_4	Disappearance and death of family member	91	J90	Pleural effusion, not elsewhere classified	97	R53_83	Other fatigue
78	T14_8	Other injury of unspecified body region	87	R91_8	Other nonspecific abnormal finding of lung field	97	K56_609	Unspecified intestinal obstruction, unspecified as to partial versus complete obstruction
78	R09_89	Other specified symptoms and signs involving the circulatory and respiratory systems	84	R47_9	Unspecified speech disturbances	96	R91_8	Other nonspecific abnormal finding of lung field
77	R49_0	Dysphonia	83	H92_10	Otorrhea, unspecified ear	91	R06_89	Other abnormalities of breathing
72	R13_10	Dysphagia, unspecified	79	R68_89	Other general symptoms and signs	85	R51	Headache
67	K21_9	Gastro-esophageal reflux disease without esophagitis	75	R06_89	Other abnormalities of breathing	82	T88_7	Unspecified adverse effect of drug or medicament
67	R05	Cough	75	J34_89	Other specified disorders of nose and nasal sinuses	81	I10	Essential (primary) hypertension
63	H91_90	Unspecified hearing loss, unspecified ear	73	H93_19	Tinnitus, unspecified ear	81	R47_89	Other speech disturbances
		....			....			....

TABLE 6. The associative relationship of the medical term-sets among the referrals in each category.

Category 1		Category 2		Category 3	
Support	Itemsets	Support	Itemsets	Support	Itemsets
0.014118	(B19_10)	0.015713	(B18_2)	0.011944	(B19_20)
0.015294	(B19_20)	0.012346	(B34_9)	0.018458	(B34_9)
0.048235	(B20)	0.015713	(B96_89)	0.013029	(B96_89)
0.010588	(B34_9)	0.066218	(B99_9)	0.074919	(B99_9)
0.016471	(B35_1)	0.014590	(C09_9)	0.021716	(C09_9)
...		...		...	
0.034118	(Z86_73, T14_90, T63_301, E11_9, K52_9, T14_8, B20)	0.010101	(H91_90, R41_82, R47_89, R47_9, S09_302)	0.011944	(R50_9, R53_81, R63_8, J03_90, T78_40)
0.034118	(Z86_73, T14_90, T63_301, E11_9, R52, T14_8, B20)	0.011223	(H91_90, R62_50, R41_82, R47_9, S09_302)	0.016287	(R09_81, J30_9, J34_89, K56_609, T78_40)
0.034118	(Z86_73, T14_90, T63_301, K52_9, R52, T14_8, B20)	0.013468	(N73_9, C18_9, I25_2, T14_90, B96_89, B99_9)	0.010858	(M41_9, R91_8, J34_89, J32_9, K63_5)
0.036471	(Z86_73, T14_90, T63_301, E11_9, R52, K52_9, T14_8)	0.010101	(H91_90, H66_90, R41_82, R47_89, R47_9, S09_302)	0.013029	(N73_9, C18_9, I25_2, T14_90, B96_89, B99_9)
0.034118	(Z86_73, T14_90, T63_301, E11_9, R52, K52_9, T14_8, B20)	0.010101	(H91_90, R62_50, H66_90, R41_82, R47_9, S09_302)	0.016287	(G47_30, G47_9, R06_83, K56_609, G47_33, T78_40)

categories. Referring to Fig. 6, both J34.89 and H91.90 have clear distinctions to be the most common terms in both categories 2 and 3. But in category 1, there are several significant highs occurring terms with R52, L98.9, R91.8, R59.0, I25.10, I10, J34.89 and E11.9 sharing the top spots. This can help the clinicians to appreciate the landscape of the referrals' group of terms that constitutes their triaged groups and may help to adjust gradually in their urgencies as the policy change. Table 5 shows the list of those terms' ICD10 codes together with their occurrence among the dataset and their respective descriptions. The next set of results is a small sample of the frequent pattern of medical term-sets discovered from the dataset using the algorithm as shown

in Table 6. It is sorted in ascending order based on their length and have min support of 0.01 under each triaged category. We can see that some of the terms did exist across the three categories but most of them have unique term-set associations. In category 1, the longest-term sets have many similar terms, and their support is above 0.03. Such patterns can be also observed in category 2, but their support values are lower. However, for category 3, there are more variations among the longer-term sets and have lower support values. The term sets' confidence versus support under each category is plotted as shown in Fig. 7. Category 2 and 3's support value is generally low below 0.025, which signifies a lower frequent pattern of occurrence in the dataset. But
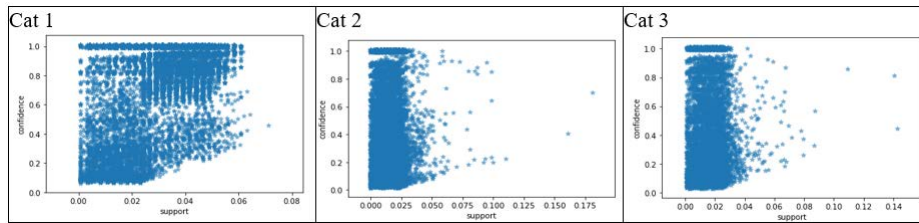
FIGURE 7. Referral's term-sets support vs confidence graphs.

TABLE 7. Term sets' antecedents-consequents relationship order according to lift value.

	Antecedents	Consequents	Antecedent support	Consequent support	Support	Confidence	Lift
			Category 1				
689	(Z63_4)	(L98_9)	0.096471	0.148235	0.011765	0.121951	0.822687
688	(L98_9)	(Z63_4)	0.148235	0.096471	0.011765	0.079365	0.822687
949	(Z63_4)	(R59_0)	0.096471	0.131765	0.010588	0.109756	0.832970
948	(R59_0)	(Z63_4)	0.131765	0.096471	0.010588	0.080357	0.832970
947	(T78_40)	(R59_0)	0.104706	0.131765	0.011765	0.112360	0.852729
...	...	...	...	...	...	...	...
3449	(I25_2, C18_9)	(N73_9, B99_9)	0.025882	0.023529	0.023529	0.909091	38.636364
4313	(R59_0, K13_70)	(J43_9, M89_9)	0.020000	0.014118	0.011765	0.588235	41.666667
4316	(J43_9, M89_9)	(R59_0, K13_70)	0.014118	0.020000	0.011765	0.833333	41.666667
4312	(R59_0, J43_9)	(K13_70, M89_9)	0.021176	0.012941	0.011765	0.555556	42.929293
4317	(K13_70, M89_9)	(R59_0, J43_9)	0.012941	0.021176	0.011765	0.909091	42.929293
			Category 2				
167	(H60_93)	(H91_90)	0.038159	0.435466	0.013468	0.352941	0.810491
166	(H91_90)	(H60_93)	0.435466	0.038159	0.013468	0.030928	0.810491
1005	(T78_40)	(S09_302)	0.124579	0.121212	0.012346	0.099099	0.817568
1004	(S09_302)	(T78_40)	0.121212	0.124579	0.012346	0.101852	0.817568
94	(H91_90)	(G47_30)	0.435466	0.028058	0.010101	0.023196	0.826701
...	...	...	...	...	...	...	...
7319	(N73_9, B99_9, T14_90)	(I25_2, B96_89, C18_9)	0.013468	0.015713	0.013468	1.000000	63.642857
7311	(N73_9, I25_2, C18_9)	(B99_9, B96_89, T14_90)	0.013468	0.015713	0.013468	1.000000	63.642857
1233	(H92_10, Z90_710)	(G45_9)	0.010101	0.015713	0.010101	1.000000	63.642857
3703	(H91_90, G45_9)	(H92_10, Z90_710)	0.014590	0.010101	0.010101	0.692308	68.538462
3706	(H92_10, Z90_710)	(H91_90, 45_9)	0.010101	0.014590	0.010101	1.000000	68.538462
			Category 3				
209	(G47_33)	(J34_89)	0.161781	0.292074	0.038002	0.234899	0.804246
208	(J34_89)	(G47_33)	0.292074	0.161781	0.038002	0.130112	0.804246
216	(J45_909)	(G47_33)	0.107492	0.161781	0.014115	0.131313	0.811674
217	(G47_33)	(J45_909)	0.161781	0.107492	0.014115	0.087248	0.811674
820	(J34_89)	(R42)	0.292074	0.057546	0.014115	0.048327	0.839798
...	...	...	...	...	...	...	...
5163	(N73_9, B99_9, C18_9)	(B96_89, T14_90)	0.013029	0.013029	0.013029	1.000000	76.750000
3722	(B96_89)	(B99_9, C18_9, T14_90)	0.013029	0.013029	0.013029	1.000000	76.750000
3724	(N73_9, B99_9, I25_2)	(B96_89)	0.013029	0.013029	0.013029	1.000000	76.750000
5154	(B96_89)	(B99_9, T14_90, I25_2, C18_9)	0.013029	0.013029	0.013029	1.000000	76.750000
5145	(I25_2, T14_90)	(B99_9, B96_89, C18_9)	0.013029	0.013029	0.013029	1.000000	76.750000

for Category 1, the rules' support value disperses up to 0.06, which signifies that there are more term sets with a higher frequency of presence. Their confidence values varied uniformly throughout the normalized range. The term sets have ranges of inter-dependencies among their items, which means the return rate varies with no concentration to specific term-set rules. Both the antecedent and consequent's support indicate the terms or term sets within the dataset, while the other three values indicate the strength of their relationships.

a) Category 1 referral

b) Category 2 referral

c) Category 3
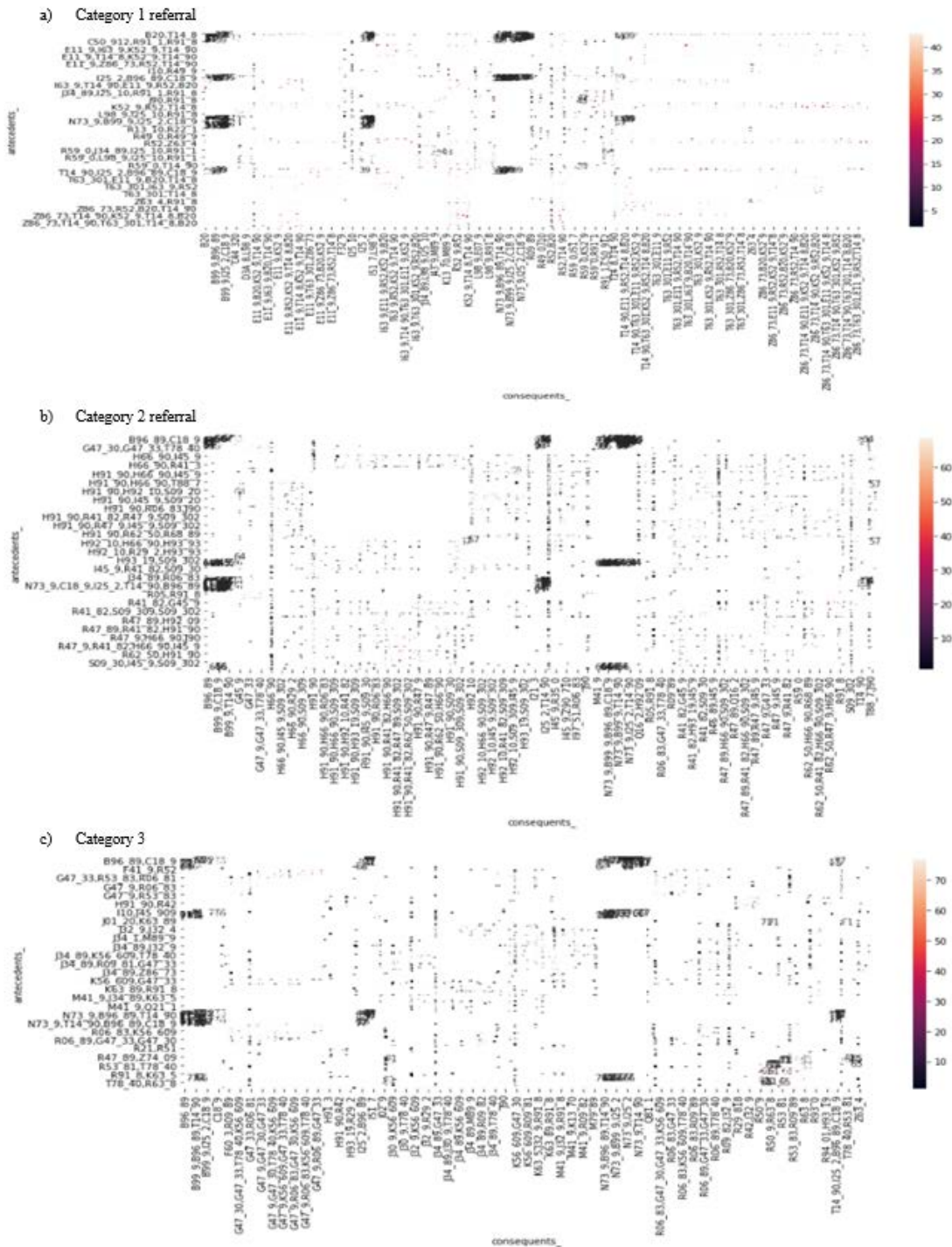
FIGURE 8. Heatmap of referral term- sets' antecedents vs consequences relationships for the three referral categories.

The details in Table 7 can be visualized as a heatmap as shown in Fig. 8. The three charts show several concentration areas where some of the term sets' relationships have a higher number of occurrences. This can assist the clinicians to appreciate the landscape of the term sets' relationship across the entire referral's dataset. Due to the large volume of data, only a subset of the term sets' rules is plotted in these charts.

B. THE RESULTS OF THE AI-BASED REFERRAL CLASSIFICATION MODEL

For this section, we run the test on two groups of models, the first group is comprised of a list of machine learning models from the sci-kit-learn library, and the second group is the customized NN constructed for this experiment. We then compare and critique the results of the two groups.

TABLE 8. Models' accuracies and confusion matrices.

Model: LR					Model: SVM				
Accuracy: 0.7772215269086358					**Accuracy: 0.7809762202753442**				
Predicted	1	2	3	All	Predicted	1	2	3	All
Actual					Actual				
1	204	18	13	235	1	215	10	10	235
2	36	204	42	282	2	52	195	35	282
3	33	36	213	282	3	43	25	214	282
All	273	258	268	799	All	310	230	259	799

Model: XGBClassifier					Model: RandomForest				
Accuracy: 0.7759699					**Accuracy: 0.81727158**				
Predicted	1	2	3	All	Predicted	1	2	3	All
Actual					Actual				
1	204	18	13	235	1	198	18	19	235
2	46	199	37	282	2	27	221	34	282
3	40	25	217	282	3	23	25	234	282
All	290	242	267	799	All	248	264	287	799

Model: SGDClassifier					Model: GaussianNB				
Accuracy: 0.7934918					**Accuracy: 0.5819774718**				
Predicted	1	2	3	All	Predicted	1	2	3	All
Actual					Actual				
1	198	15	22	235	1	105	119	11	235
2	32	215	35	282	2	14	252	16	282
3	34	27	221	282	3	37	137	108	282
All	264	257	278	799	All	156	508	135	799

Model: KNeighborsClassifier					Model: MLPClassifier-sgd				
Accuracy: 0.6658322903					**Accuracy: 0.72590738423**				
Predicted	1	2	3	All	Predicted	1	2	3	All
Actual					Actual				
1	189	26	20	235	1	195	21	19	235
2	57	180	45	282	2	52	182	48	282
3	58	61	163	282	3	42	37	203	282
All	304	267	228	799	All	289	240	270	799

Model: MLPClassifier-lbfgs					Model: MLPClassifier-adam				
Accuracy: 0.8147684605					**Accuracy: 0.82227784730**				
Predicted	1	2	3	All	Predicted	1	2	3	All
Actual					Actual				
1	192	14	29	235	1	199	17	19	235
2	28	221	33	282	2	26	229	27	282
3	22	22	238	282	3	28	25	229	282
All	242	257	300	799	All	253	271	275	799

1) RESULTS FROM GROUP 1

For the first round of tests, we use a series of statistical and machine learning models: linear regression (LR), support vector machine (SVM), Random forest, Extreme Gradient Boosting Classifier (XGBclassifer), Gaussian Naive Bayes (GaussianNB), KNeighbours-classifier, and three Multi-level Perception (MLP) classifiers with different learning strategies such as Stochastic Gradient descent (SGD), Adam, and limited-memory Broyden, Fletcher, Goldfarb, and Shanno (lbfgs). Their accuracy and their confusion matrix results are tabulated in Table 8. Though the hyperparameters of the models have not been changed, the majority of the model's accuracy is from 58% to 82%. A confusion matrix is generated for each model to determine and compare the numbers of actual and predicted samples' labels.

The result shows that there are significant error gaps. While we observe the need not to over-fit the models, the error values are unacceptable. As this is related to the medical term, having higher accuracy in classification is a critical priority.

All the models' accuracy are compiled and plotted in Fig. 9, with the MLP models and the Randomforests managed to achieve accuracy above 80%. The others fair with the range of 70-80% except for KNeighbours-Classifier and GaussianNB which score less than 70%. The results can be improved if their hyperparameters can be tuned, however, our goal is to use deep learning to build a customized NN model for the text classification.

2) RESULTS FROM GROUP 2

In the second group of tests, we use the neural network built based on the specification set in the previous section. Both the training and testing dataset have been vectorized. The model has been trained and tested with the dataset using the ratio of 80:20 and 90:10 split.
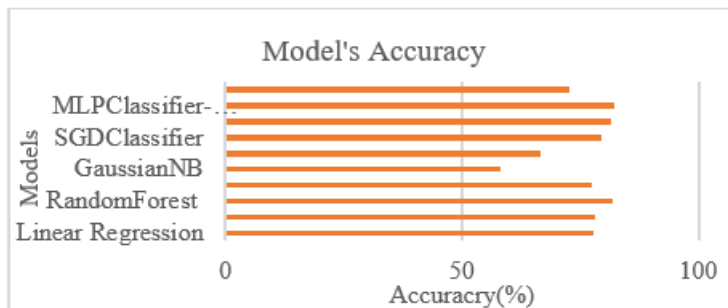
FIGURE 9. Various model's training and validation results using referral datasets.
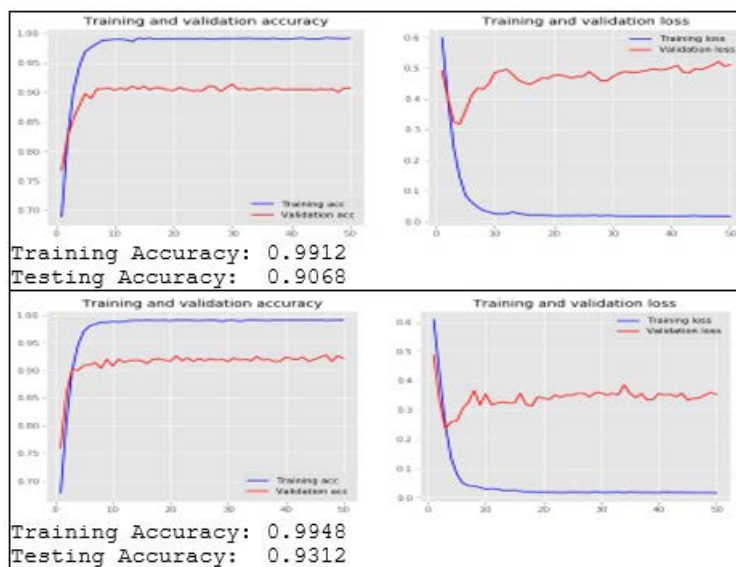
FIGURE 10. Customized neural network results using 80:20 and 90:10 dataset split.

The results are shown in Fig. 10. As the first result shows the model that has been trained with an 80:20 split has an accuracy of 90% with validation loss at 0.5, the second test that uses 90:10 achieved a higher accuracy of 93% and lower validation loss of <0.4. This can be attributed to a better-trained model which has been achieved with more training datasets.

3) COMPARISONS OF RESULTS FROM GROUP 1 AND 2

We pit the test samples against the model and create a confusion matrix between the actual test labels against the predictions as shown in Fig. 11. Compared to the current model's confusion results against those of the previous models' results in Group 1, we can see that there is a big improvement in the accuracy.

Referring to Table 9, only the customized NN developed for this project managed to achieve higher effectiveness results as compared to the rest, scoring above 90% in terms of accuracy. Based on the computational time taken for the algorithms to complete, LR, SVM, Randomforest and XGB classifier took on an average of 5 seconds to complete while

the MLP models require twice as long to complete. The customized NN that is used for the final classification took the longest with the score of over 20 minutes due to the NN's depth, with high value set for both the epoch and iterations which was set to 500 and 2000. Accuracy is paramount in health and medicine settings. Although the running time of customized NN is relatively large, we have achieved our research goal of developing model that can achieve the highest accuracy. Currently, the models were run on a desktop computer with the configuration of Intel i7 CPU, 8GB RAM and 500GB HDD. To compensate the current deficiency and for further enhancement the parallel computing techniques can be applied and more powerful computing devices can be used.

C. THE RESULTS OF OUTLIER DETECTION FROM AI-BASED REFERRAL INSIGHT MODEL

The technique is applied against the referrals under their respective categories. Referring to the following diagrams and tables, the vectorized referrals' contents are visualized while the outliers have been identified and reversed back

```
Predicted   0    1    2    3    All
Actual
1           0    76   6    2    84
2           0    7    72   6    85
3           1    4    2    91   98
All         1    87   80   99   267
```
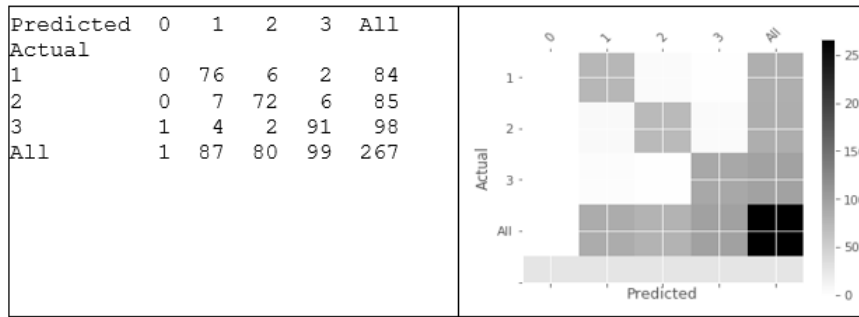
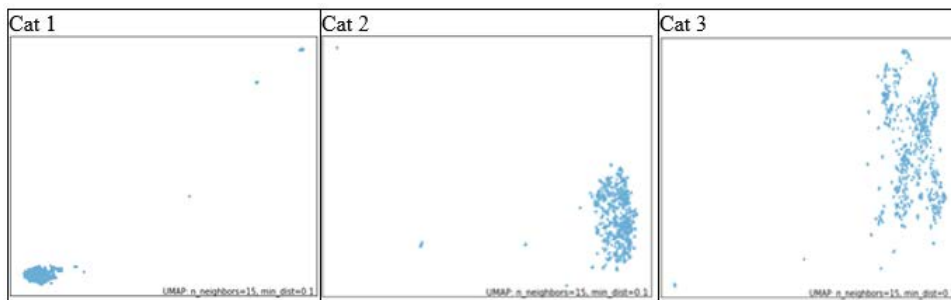**FIGURE 11.** Confusion matrix results for the final model.

FIGURE 12. Scatter plot of UMAP embedding with category 1, 2, 3 referrals.

TABLE 9. Different model's performance and execution time.

Model	Execution time (second)	Accuracy
LR	1.93	0.78
SVM	6.53	0.78
RandomForest	5.91	0.82
XGBClassifier	4.46	0.78
GaussianNB	5.33	0.58
KNeighbors-Classifier	5.19	0.67
SGDClassifier	12.35	0.79
MLPClassifier-lbfgs	23.87	0.81
MLPClassifier-adam	24.73	0.82
MLPClassifier-sgd	26.12	0.73
Customized NN	1389	0.93

FIGURE 13. Scatter plot of all three categories referrals combined.

to their original contents with ICD10 codes. The level of contamination parameter allows the user to change the sensitivity of the outlier computation, with a lower value to increase the LOF's stringency and produce fewer outliers, and vice versa. All the scatter plots as shown in Fig. 12, and Fig. 13 is scaled based on their range limits.

Fig. 12 showed that the differences between the outlying referrals and the rest under category 1 are high, whereas, for those under category 2, the spread is less as compared with those in category 1. While those under category 2 have a bigger spread but not extreme. The outliers among the 3 categories can be identified using the LOF and their outputs have transformed back to the ICD10 codes as shown in Table 10, based on the contamination hyperparameter value of 0.005.

This value is selected to demonstrate the ability of LOF in extracting out a small set of outliers. This can be changed

to suit the users' requirements if they want to see more outlying referrals. For those under category 1, the top few have the common ICD10 representation starting with the first combination of B20, C34.90, D49.1, E11_9 and I49.9. Those in cat 3 have similarities among themselves on the ICD10 codes of B96.89, B99.9, C18.9. but those under cat 2 don't share such patterns. All their ICD10 descriptions are tabulated in Table 11 under their respective categories. The outliers here represent the uniqueness of the referrals which don't normally conform to the usual clusters and commonalities of referrals, rather than erroneous types. These records may be of some interest to the triaging clinicians, who wish to investigate further.

TABLE 10. Outliers' referrals medical terms for category 1, 2, 3.

Category 1
array(['c34_90','d49_1','e07_9','i49_9','i65_23','j04_30','j32_9','j34_1','j34_89','j43_9','j98_4','k13_70','k63_89','m51_26','m62_838','m89 9','r17','r59 0','r91 1','t14 8'],
array(['b20','c34_90','d49_1','e11_9','i49_9','i65_23','j32_9','j34_1','j43_9','j98_4','k13_70','k52_9','k63_89','m62_838','m89_9','r13_10','r46 89','r52','r59 0','r91 1','t14 8','t14 90','t63 301','z86 73'],
array(['b20','c34_90','d49_1','e11_9','i49_9','i65_23','j32_9','j34_1','j43_9','j98_4','k13_70','k52_9','k63_89','m62_838','m89_9','r13_10','r46 89','r52','r59 0','r91 1','t14 8','t14 90','t63 301','z86 73'],
array(['b20','c34_90','d49_1','e11_9','i49_9','i65_23','j32_9','j34_1','j43_9','j98_4','k13_70','k52_9','k63_89','m62_838','m89_9','r13_10','r46 89','r52','r59 0','r91 1','t14 8','t14 90','t63 301','z86 73'],
array(['b20','c34_90','d49_1','e11_9','i49_9','i65_23','j32_9','j34_1','j43_9','j98_4','k13_70','k52_9','k63_89','m62_838','m89_9','r13_10','r46 89','r52','r59 0','r91 1','t14 8','t14 90','t63 301','z86 73']]
Category 2
array(['e53_8','g45_9','g47_9','h66_90','h91_90','h92_10','h93_19','h93_213','i45_9','i97_51','j00','j11_1','k08_89','l65_9','r06_83','r35_0','r41 82','r45 86','r46 89','r51','r68 89','r94 112','s09 30','s09 302','s09 309','z90 710']),
array(['g47_9','h66_90','h91_90','h92_10','h93_213','i21_3','i45_9','i97_51','j00','j11_1','k08_409','l65_9','r06_83','r41_82','r45_89','r46_8 9','r51','r91 8','r94 112','s09 20','s09 302','s09 309']),
array(['g47_9','h66_90','h91_90','h92_10','h93_213','i21_3','i45_9','i97_51','j00','j11_1','k08_409','l65_9','r06_83','r41_82','r45_89','r46_8 9','r51','r91 8','r94 112','s09 20','s09 302','s09 309']),
array(['d80_3','h73_893','h91_90','i25_82','i69_328','r01_2','r11_0','s09 30','s09_302']),
array(['e53_8','g45_9','g47_9','h66_90','h91_90','h92_10','h93_19','h93_213','i45_9','i97_51','j00','j11_1','k08_89','l65_9','r06_83','r35_0','r41 82','r45 86','r46 89','r51','r68 89','r94 112','s09 30','s09 302','s09 309','z90 710'])]
Category 3
array(['b96_89','b99_9','c18_9','d49_1','g20','i25_10','i25_2','m06_9','m35_9','n73_9','r69','t14_90'],
array(['b96 89','b99 9','c18 9','h91 90','h93 19','i25 10','i25 2','n73 9','r42','t14 90','z77 098'],
array(['b96_89','b99 9','c18_9','d50_9','i25_2','n73_9','r69','t14_90'],
array(['b96 89','b99 9','c18 9','e23 0','i25 2','l60 8','n73 9','r63 4','t14 90'],
array(['b96 89','b99 9','c18 9','i25 2','n73 9','r69','t14 90'])]

TABLE 11. Common outliers' ICD10 codes for three categories.

Category 1	Category 2	Category 3
C34.90- Malignant neoplasm of unspecified part of unspecified bronchus or lung D49.1 - Neoplasm of unspecified behaviour of respiratory system E07.9 - Disorder of thyroid, unspecified I49.9 - Cardiac arrhythmia, unspecified I65.23- Occlusion and stenosis of bilateral carotid arteries	E53.8 - Deficiency of other specified B group vitamins G45.9 - Transient cerebral ischemic attack, unspecified G47.9 - Sleep disorder, unspecified H66.90- Otitis media, unspecified, unspecified ear H91.90- Unspecified hearing loss, unspecified ear G47.9 - Sleep disorder, unspecified D80.3 - Selective deficiency of immunoglobulin G [IgG] subclasses E53.8 - Deficiency of other specified B group vitamins G45.9 - Transient cerebral ischemic attack, unspecified	B96.89 - Other specified bacterial agents as the cause of diseases classified elsewhere B99.9 - Unspecified infectious disease C18.9 - Malignant neoplasm of colon, unspecified' respectively

The final scatter plot as shown in Fig. 13 has all the 3 referrals consolidated. Their scale is mapped to their range limits, so the plot doesn't correlate visually with the previous Fig. 12. Their dimensionalities are reduced accordingly to the entire referral dataset, so their values are relative to the collective whole.

VI. CONCLUSION AND FUTURE WORK

This study used the combination of a Medical NLP and association rules mining to acquire insights into GP referrals to otorhinolaryngology. The referrals were originally in digital document form and are unstructured free texts. The RAII performed data preparation and converted them to digital form, followed by using the medical NLP to extract the medical term features. It then selected only those terms with the highest score and was organised into a list of medical terms that were based on ICD10 codes. Association rules mining was applied against the dataset, extracting multiple associative rules among the medical term sets across all the parsed referrals. The results showed that this approach can produce significant insight into the referrals' datasets with ease. This information is particularly valuable for clinicians and healthcare management who need to know the landscape of the referrals from a holistic perspective. In analysing this data, it may be possible to target resource distribution to vulnerable patients and those with the highest level of need.

The neural network driven triaging approach has shown that it achieved high accuracy results in classifying the

referrals based on the test dataset which uses the medical-based NLP system to derive the relevant medical codes from the GP referrals. Prior to text classification, the text data needed to be pre-processed with both stemming and lemmatization, which are normalization techniques that are used before they can be further processed. However, these pre-processing of NLPs was not suitable to handle the meaning of words within the medical field where each term has a critical relationship to the well-being of the patients. Therefore, a specialized NLP developed for the healthcare sector should be used for this intent. We consider that deriving out the medical diagnosis of the referrals' content into clinical codes such as ICD10CM and RxNorm will promote the better accuracy and reliability of the referral's text classification process. Our experiment showed that our approach can achieve a significantly higher accuracy result as compared to the results of the research in [28]. This is our initial experiment, and we expect future enhancements to include exogenous conditions such as social and demographic considerations. Besides, high-risk categories such as comorbidity or pre-requisite for other medical attention can be used to fine-tune the triaging process even further.

For the approach to locate the referrals' outliers for the three categories, it uses the ensemble of the dimensionality reduction algorithm and the outlier detection, on the processed data that had been processed with the aid of the Medical NLP, including the vectorization process. The approach can then extract the array of features that are regarded as abnormal based on the density's threshold of the LOF, and that in term can be referred to the specific medical referrals.

The level of outlier's detection can be adjusted as required to support the clinical process improvement. This can lead to better management of the clinician's triaging process and improve their quality of work.

AUTHOR CONTRIBUTIONS

Chee Keong Wee and Xujuan Zhou have proposed the idea and consulted it with Raj Gururajan, Xiaohui Tao, Jennifer Chen, Rashmi Gururajan, Nathan Wee, and Prabal D. Barua, and all authors contributed to the development and completion of the idea, Chee Keong Wee conducted the main part of the algorithms design and experiments, and all authors participated in discussions and structure of the articles, and writing the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

[1] J. Todd, B. Richards, B. Vanstone, and A. Gepp, "Text mining and automation for processing of patient referrals," *Appl. Clin. Informat.*, vol. 9, no. 1, pp. 232–237, Jan. 2018.

[2] Queensland Health. *Smart Referrals 2021*. Accessed: Jun. 10, 2021. [Online]. Available: https://www.health.qld.gov.au/sunshinecoast/healthcare-providers/referrals/smart-referrals

[3] H. Huggard, Y. S. Koh, G. Dobbie, and E. Zhang, "Detecting concept drift in medical triage," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2020, pp. 1733–1736.

[4] L. A. Stainkey, I. A. Seidl, A. J. Johnson, G. E. Tulloch, and T. Pain, "The challenge of long waiting lists: How we implemented a GP referral system for non-urgent specialist' appointments at an Australian public hospital," *BMC Health Services Res.*, vol. 10, no. 1, pp. 1–5, Dec. 2010.

[5] T. P. McVeigh, D. Donnelly, M. Al Shehhi, E. A. Jones, A. Murray, S. Wedderburn, M. Porteous, and S. A. Lynch, "Towards establishing consistency in triage in a tertiary specialty," *Eur. J. Hum. Genet.*, vol. 27, no. 4, pp. 547–555, Apr. 2019.

[6] Queensland Health. (2021). *Referral Guidelines Queensland Health*. Accessed: Jun. 10, 2021. [Online]. Available: https://www.health.qld.gov.au/sunshinecoast/healthcare-providers/referrals

[7] Queensland Health. (2021). *Clinical Prioritisation Criteria*. Accessed: Jun. 2021. [Online]. Available: https://cpc.health.qld.gov.au/

[8] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform manifold approximation and projection for dimension reduction," 2018, *arXiv:1802.03426*.

[9] M. M. Breunig, H. P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, May 2000, pp. 93–104.

[10] P. Bhatia, B. Celikkaya, M. Khalilia, and S. Senthivel, "Comprehend medical: A named entity recognition and relationship extraction web service," in *Proc. 18th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2019, pp. 1844–1851.

[11] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *Proc. Int. Conf. Eng. Technol. (ICET)*, Aug. 2017, pp. 1–6.

[12] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad, "State-of-the-art in artificial neural network applications: A survey," *Heliyon*, vol. 4, no. 11, Nov. 2018, Art. no. e00938.

[13] E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I. W. H. Kwok, L. G. Ng, F. Ginhoux, and E. W. Newell, "Dimensionality reduction for visualizing single-cell data using UMAP," *Nature Biotechnol.*, vol. 37, no. 1, pp. 38–44, Jan. 2019.

[14] H. Ma, Y. Hu, and H. Shi, "Fault detection and identification based on the neighborhood standardized local outlier factor method," *Ind. Eng. Chem. Res.*, vol. 52, no. 6, pp. 2389–2402, 2013.

[15] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, "Natural language processing: An introduction," *J. Amer. Med. Inform. Assoc.*, vol. 18, no. 5, pp. 544–551, 2011.

[16] X. Zhou, R. Gururajan, Y. Li, R. Venkataraman, X. Tao, G. Bargshady, and S. Kondalsamy-Chennakesavan, "A survey on text classification and its applications," in *Web Intelligence*, vol. 18, no. 3. Amsterdam, The Netherlands IOS Press, 2002, pp. 205–216.

[17] (2021). *How Amazon Comprehend Medical Works Amazon Comprehend*. Accessed: Jun. 10, 2021. [Online]. Available: https://docs.aws.amazon.com/comprehend-medical/latest/dev/comprehendmedical-howitworks.html

[18] N. W. Sterling, R. E. Patzer, M. Di, and J. D. Schrager, "Prediction of emergency department patient disposition based on natural language processing of triage notes," *Int. J. Med. Informat.*, vol. 129, pp. 184–188, Sep. 2019.

[19] N. Roysden and A. Wright, "Predicting health care utilization after behavioral health referral using natural language processing and machine learning," in *Proc. AMIA Annu. Symp.*, 2015, p. 2063.

[20] Y. Juhn and H. Liu, "Artificial intelligence approaches using natural language processing to advance EHR-based clinical research," *J. Allergy Clin. Immunol.*, vol. 145, no. 2, pp. 463–469, Feb. 2020.

[21] Z. Wang, P. A. Keane, M. Chiang, C. Y. Cheung, T. Y. Wong, and D. S. W. Ting, "Artificial intelligence and deep learning in ophthalmology," in *Artificial Intelligence in Medicine*. Amsterdam, The Netherlands: Elsevier, 2020, pp. 1–34.

[22] J. He, S. L. Baxter, J. Xu, J. Xu, X. Zhou, and K. Zhang, "The practical implementation of artificial intelligence technologies in medicine," *Nature Med.*, vol. 25, no. 1, pp. 30–36, Jan. 2019.

[23] T. Davenport and R. Kalakota, "The potential for artificial intelligence in healthcare," *Future Healthcare J.*, vol. 6, no. 2, pp. 94–98, Jun. 2019.

[24] W. S. Hong, A. D. Haimovich, and R. A. Taylor, "Predicting hospital admission at emergency department triage using machine learning," *PLoS ONE*, vol. 13, no. 7, Jul. 2018, Art. no. e0201016.

[25] Y. Raita, T. Goto, M. K. Faridi, D. F. M. Brown, C. A. Camargo, and K. Hasegawa, "Emergency department triage prediction of clinical outcomes using machine learning models," *Crit. Care*, vol. 23, no. 1, pp. 1–13, Dec. 2019.

[26] T. Goto, C. A. Camargo, M. K. Faridi, R. J. Freishtat, and K. Hasegawa, "Machine learning–based prediction of clinical outcomes for children during emergency department triage," *JAMA Netw. Open*, vol. 2, no. 1, Jan. 2019, Art. no. e186937.

[27] S. Horng, D. A. Sontag, Y. Halpern, Y. Jernite, N. I. Shapiro, and L. A. Nathanson, "Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning," *PLoS ONE*, vol. 12, no. 4, Apr. 2017, Art. no. e0174708.

[28] Y. Tan, S. Bacchi, R. J. Casson, D. Selva, and W. Chan, "Triaging ophthalmology outpatient referrals with machine learning: A pilot study," *Clin. Experim. Ophthalmol.*, vol. 48, no. 2, pp. 169–173, Mar. 2020.

[29] S. Swaminathan, K. Qirko, T. Smith, E. Corcoran, N. G. Wysham, G. Bazaz, G. Kappel, and A. N. Gerber, "A machine learning approach to triaging patients with chronic obstructive pulmonary disease," *PLoS ONE*, vol. 12, no. 11, Nov. 2017, Art. no. e0188532.

[30] S. W. Choi, T. Ko, K. J. Hong, and K. H. Kim, "Machine learning-based prediction of Korean triage and acuity scale level in emergency department patients," *Healthcare Inform. Res.*, vol. 25, no. 4, pp. 305–312, 2019.

[31] B. W. Patterson, C. J. Engstrom, V. Sah, M. A. Smith, E. A. Mendonça, M. S. Pulia, M. D. Repplinger, A. G. Hamedani, D. Page, and M. N. Shah, "Training and interpreting machine learning algorithms to evaluate fall risk after emergency department visits," *Med. Care*, vol. 57, no. 7, pp. 560–566, Jul. 2019.

[32] A. T. Elder, I. C. Mcmanus, A. Patrick, K. Nair, L. Vaughan, and J. Dacre, "The value of the physical examination in clinical practice: An international survey," *Clin. Med.*, vol. 17, no. 6, pp. 490–498, Dec. 2017.

[33] J. E. Kim-Hwang, A. H. Chen, D. S. Bell, D. Guzman, H. F. Yee, and M. B. Kushel, "Evaluating electronic referrals for specialty care at a public hospital," *J. Gen. Internal Med.*, vol. 25, no. 10, pp. 1123–1128, Oct. 2010.

[34] B. Guzman, I. Metzger, Y. Aphinyanaphongs, and H. Grover, "Assessment of Amazon comprehend medical: Medication information extraction," 2020, *arXiv:2002.00481*.

[35] I. Spasic and K. Button, "Patient triage by topic modeling of referral letters: Feasibility study," *JMIR Med. Informat.*, vol. 8, no. 11, Nov. 2020, Art. no. e21252.

[36] K. Lau, J. Wilkinson, and R. Moorthy, "A web-based prediction score for head and neck cancer referrals," *Clin. Otolaryngol.*, vol. 43, no. 4, pp. 1043–1049, Aug. 2018.

[37] M. Abdar, M. Zomorodi-Moghadam, X. Zhou, R. Gururajan, X. Tao, P. D. Barua, and R. Gururajan, "A new nested ensemble technique for automated diagnosis of breast cancer," *Pattern Recognit. Lett.*, vol. 132, pp. 123–131, Apr. 2020.

[38] C. Kok, V. Jahmunah, S. L. Oh, X. Zhou, R. Gururajan, X. Tao, K. H. Cheong, R. Gururajan, F. Molinari, and U. R. Acharya, "Automated prediction of sepsis using temporal convolutional network," *Comput. Biol. Med.*, vol. 127, Dec. 2020, Art. no. 103957.

[39] G. Bargshady, J. Soar, X. Zhou, R. C. Deo, F. Whittaker, and H. Wang, "A joint deep neural network model for pain recognition from face," in *Proc. IEEE 4th Int. Conf. Comput. Commun. Syst. (ICCCS)*, Feb. 2019, pp. 52–56.

[40] G. Bargshady, X. Zhou, R. C. Deo, J. Soar, F. Whittaker, and H. Wang, "Ensemble neural network approach detecting pain intensity from facial expressions," *Artif. Intell. Med.*, vol. 109, Sep. 2020, Art. no. 101954.

[41] E. Nasarian, M. Abdar, M. A. Fahami, R. Alizadehsani, S. Hussain, M. E. Basiri, M. Zomorodi-Moghadam, X. Zhou, P. Pławiak, U. R. Acharya, R.-S. Tan, and N. Sarrafzadegan, "Association between work-related features and coronary artery disease: A heterogeneous hybrid feature selection integrated with balancing approach," *Pattern Recognit. Lett.*, vol. 133, pp. 33–40, May 2020.

[42] M. Zomorodi-Moghadam, M. Abdar, Z. Davarzani, X. Zhou, P. Pławiak, and U. R. Acharya, "Hybrid particle swarm optimization for rule discovery in the diagnosis of coronary artery disease," *Expert Syst.*, vol. 38, no. 1, Jan. 2021, Art. no. e12485.

[43] M. Abdar, E. Nasarian, X. Zhou, G. Bargshady, V. N. Wijayaningrum, and S. Hussain, "Performance improvement of decision trees for diagnosis of coronary artery disease using multi filtering approach," in *Proc. IEEE 4th Int. Conf. Comput. Commun. Syst. (ICCCS)*, Feb. 2019, pp. 26–30.

CHEE KEONG WEE received the Ph.D. degree in artificial intelligence on part-time basis from the University of Southern Queensland (USQ), in April 2022. He is an experienced IT professional with over 26 years of IT work experience, and he is currently working as a Data Engineer/Scientist with eHealth Queensland, Australia. His research interests include healthcare AI covering prediction, natural language processing, and deep reinforcement learning.

XUJUAN ZHOU (Member, IEEE) received the Ph.D. degree from the Queensland University of Technology, Australia. She is currently an Associate Professor with the University of Southern Queensland (USQ), Australia. Her research interests include machine learning and deep learning, health informatics, information retrieval/filtering, recommender systems, sentiment analysis, and natural language processing. In 2021, she received a three-year (2022–2024) Advance Queensland Industry Research Fellowship grant for the project titled "Early detection of chronic health conditions using AI prediction model." She has completed the Commonwealth Innovation Connections Grant research project titled "Prediction of cancer recurrence using innovative machine learning approaches" (2017–2019). She published papers in prestige journals (e.g., *Artificial Intelligence Review*, *Decision Support Systems*, *Pattern Recognition Letter*, *Artificial Intelligence in Medicine*, *Journal of Medical Internet Research*, *Knowledge-Based Systems*, and IEEE ACCESS). She is an Editorial Board Member for *Web Intelligence*—an international journal and a reviewer for many journals and conferences related to her research fields.

RAJ GURURAJAN is currently a Professor with the University of Southern Queensland (USQ), Australia. His research record to date includes over 300 refereed publications with three best paper awards, over six millions Australian dollar in total research grants, completion of 59 research theses, Queensland Premier's Research Excellence Award for the Patient Journey Board project, USQ Research Excellence Award, and an Endeavour Executive Award. He has delivered many keynote addresses in health informatics, including the prestigious SAARCC event. In addition, he is currently the Treasurer of the Australian Committee of Professors and Heads of Information Systems (ACPHIS). He has also been invited to accompany the Queensland Minister for Science and Technology on international delegations, including the Boston Biotech event, in 2012. In 2009, he led a delegation to India to explore the Australia-India Obesity Institute, however this did not materialize due to change in governments.

XIAOHUI TAO (Senior Member, IEEE) is currently an Active Researcher in AI and data science and an Associate Professor (computing) with the School of Mathematics, Physics and Computing, University of Southern Queensland (USQ), Australia. His research outcomes have been published on many top-tier journals (e.g., IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, *IPM*, *KBS*, and *ESWA*) and conferences (e.g., ICDE, CIKM, PAKDD, and WISE). His research interests include data analytics, natural language processing, machine learning, knowledge engineering, information retrieval, and health informatics. He is a Senior Member of ACM. He received the ARC DP Grant (2022–2024) and an Australian Endeavour Research Fellow (2015–2016), along with many other schemes.

JENNIFER CHEN currently practices as an Otolaryngology Unaccredited at the Royal Brisbane and Women's Hospital and is currently an Associate Academic Lecturer with The University of Queensland. She has a keen interest in the field of head and neck cancer and is aiming to complete her Masters of Philosophy thesis on second primary malignancy in head and neck cancer surveillance by the end of this year.

NATHAN WEE graduated from The University of Queensland, in 2021. He is an IT professional working at Dialog, as a Data Analyst, and he supports his company's corporate client, such as Queensland Health in various large-scale IT projects. He has acquired various data science and IT certifications with Coursera and Microsoft. His research interest includes the application of AI for healthcare.

RASHMI GURURAJAN is a final year Psychiatry Registrar at the Royal Brisbane and Women's Hospital. She is specializing in the field of Consultation-Liaison Psychiatry and has special interests in psycho-oncology and palliative care, pain medicine, and the integration of artificial intelligence technology with clinical medicine. In particular, she hopes to be able to use AI techniques to optimize the delivery of health services and to optimize resource distribution.

PRABAL DATTA BARUA received the Ph.D. degree in information system from the University of Southern Queensland. He is currently an Adjunct Professor with the University of Southern Queensland and an Honorary Industry Fellow at the University of Technology Sydney. He is an academic and accredited research supervisor at the University of Southern Queensland with 15 years of teaching experience. He received research support from the Queensland Government Innovation Connections under the Entrepreneurs program to research ''Cancer recurrence using innovative machine learning approaches.'' He is interested in AI technology development in health, education, agriculture, and environmental science, and he published several papers in the Q1 journal. He is the industry leader in ICT entrepreneurship in Australia and sitting as an ICT advisory panel member of many organizations.

● ● ●