Active Learning Strategy for COVID-19 Annotated Dataset

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 19, 2021, accepted November 21, 2021, date of publication November 23, 2021,
date of current version December 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3130383

Active Learning Strategy for COVID-19 Annotated Dataset

AMRIL NAZIR[1] AND RICKY MAULANA FAJRI[2,3]

[1]Department of Information Systems, College of Technological Innovation, Abu Dhabi Campus, Zayed University, Abu Dhabi, United Arab Emirates
[2]Department of Mathematics and Computer Science, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands
[3]Department of Information Technology, Indo Global Mandiri University, Palembang 30129, Indonesia

Corresponding author: Amril Nazir (mohd.nazir@zu.ac.ae)

ABSTRACT The efficient diagnosis of COVID-19 plays a key role in preventing its spread. Recently, many artificial intelligence techniques, such as the deep neural network approach, have been implemented to help efficient diagnosis of COVID-19. However, the accurate performance of deep learning depends on the tuning of many hyperparameters and a large amount of labeled data. This COVID-19 data bottleneck also leads to insufficient human resources for data labeling, which presents a challenging obstacle. In this paper, a novel discriminative batch-mode active learning (DS3) is proposed to allow faster and more effective COVID-19 data annotation. The framework specifically designed to suit the imbalanced data phenomenon that is characteristic of COVID-19 data. Extensive experiments over four public real-world COVID-19 datasets from several countries such as Brazil, China, Israel and Mexico show that our active learning framework significantly outmatches other state-of-the-art models. Our proposed framework achieves an average G-Mean of 10% improvement for the four datasets. Finally, the results of significance testing verify the effectiveness of DS3 and its superiority over baseline active learning algorithms.

INDEX TERMS COVID-19, imbalanced data, active learning, deep neural network.

I. INTRODUCTION

The COVID-19 pandemic has infected over 10 million people globally with more than 4 million people deceased as of mid-June 2021. This crisis has further affected billions of people on a social, economic, and medical level, leading to significant changes in social connections, health regulations, commerce, employment, and educational settings. Thus, the pandemic is a threat to human society, and fast action is required. In reaction to this, The COVID-19 pandemic has motivated the scientific community to assist front-line medical personnel with cutting-edge research for viral mitigation, detection, and prevention. By utilizing digital technologies, the scientific community has made two significant contributions to the fight against COVID-19. The primary digital effort came from the Artificial Intelligence (AI) community in the form of automatic COVID-19 identification from Computed Tomography (CT) scans and X-ray pictures. Secondly, mathematicians and epidemiologists are creating comprehensive virus dispersion and transmission models to predict virus propagation under different mobility and social distance situations [1]. Aside from these examples, other attempts are being made to analyze social and emotional behavior from social media [2], to gather academic articles for knowledge-based discovery [3], and to identify COVID-19 from cough samples [4]. Artificial Intelligence, machine learning, and deep learning have shown promising results in solving real-world problems using image recognition [5]–[8], natural language processing [9]–[11], and speech recognition [12]–[14]. Recently, researchers have employed a deep neural network to perform automatic COVID-19 detection. In the standard process of diagnosis, a patient will undergo numerous screening tests, such as clinical assessments, laboratory tests, chest X-ray, and PCR testing to rule out pneumonia and confirm COVID-19 infection. Deep learning plays an important role in speeding up this process by providing automatic COVID-19 detection using chest X-ray recognition; it has shown a significant impact in accurate detection of COVID-19 patients [15]–[17]. Although deep learning can make an accurate prediction, the performance of deep neural networks on COVID-19 task depends on large numbers of labeled data, such as chest CT scans and symptoms. However, the process of data labelling is not only time-consuming and arduous but also requires the expertise of medical professionals. Furthermore, sufficient annotation of COVID-19 data is impossible due to the fast spread of the virus, time limitations,

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.

and the heavy burden on the healthcare system. To solve this problem, many works have implemented active learning to reduce the annotation time and effort. Active learning works by intelligently selecting the most informative samples to be labeled by a domain expert. It is expected that active learning will be able to maintain the model performance while reducing the annotation cost. Recently, many works have combined active learning with deep neural networks. The underlying idea for this combination is that deep neural networks can detect complex data representation, such as images, and thus improve the prediction outcome. Furthermore, deep active learning also improves the automatic detection of COVID-19 through lung x-ray image recognition.

However, this study has a drawback as it is challenging to implement to real human since the deep neural network retraining procedure require certain amount of time thus affecting human mental awareness [2]. Furthermore, in the real world the dataset of COVID-19 is significantly imbalanced. The number of positive cases is small compared to negative cases, making positive cases a minority class. Therefore, it is a challenge for the active learning model to select the most informative samples. In several works, informative samples are regarded as the samples that will improve the model performance [18]–[20]; however, for imbalanced data, it is also essential to select the minority class.

In this paper, a novel approach in active learning framework to reduce annotation cost is proposed. We propose a discriminative batch-mode active learning framework, called DS3[1], to implement a discriminative, skew-specialized sampling that is suitable for imbalanced data. The experimental results demonstrate that DS3 can greatly cut the annotation cost for training a model and consistently outperforms the state-of-the-art active learning methods in the diagnosis of COVID-19. The contribution of the paper can be summarized as:

- We propose a novel batch-mode active learning specifically designed to solve imbalanced data annotation.
- We perform discriminative batch-mode active learning that outperforms the state-of-the-art active learning approaches in cutting the labeling cost and achieves effective diagnosis of COVID- 19.
- We perform experiments on four real COVID-19 patient datasets. We compare the DS3 algorithm with other state-of-the-art batch-mode active learning algorithms. We statistically test the performance of our model compared to other popular classifiers with the Wilcoxon test. The results show that DS3 outperforms most other state-of-the-art batch-mode active learning model.

The remainder of this paper is organized as follows. A brief discussion of the related studies in section II. The detailed description of the experimental materials, proposed framework, and algorithms will be shown in Section V. Section VI demonstrates the experimental results and corresponding empirical evaluation. Section VII presents a discussion about this work. This paper is concluded with an assessment of future work in Section VIII.

[1]The source code and datasets of this work are publicly available at https://github.com/analyticray/Discriminative-Batch- Mode-Active-Learning-Framework-DS3-

II. RELATED WORK

COVID-19 pandemic has attracted many researchers to develop state-of-the-art models in performing automatic detection. For example Mohammed *et al.* [17] proposed a multi-criteria decision making (MCDM) to evaluate and benchmark the different diagnostic models for COVID-19 with respect to evaluation criteria. Another study by Al-Waisy *et al.* [15] proposed a novel multimodal deep learning system for identifying COVID-19 data based on X-Ray Images. The proposed DeepNet architecture showed promising result in COVID-19 prediction. Similarly, Al-Waisy *et al.* [16] proposed an advanced ResNet34 deep neural network image recognition model to classify healthy and COVID-19 infected patient based on the X-Ray images. More recently, Mohammed *et al.* [21] performed a large comparison study on various machine learning and deep learning models. The study reported that ResNet50 achieved the optimum accuracy of 98.8%. Although most of these studies showing promising results, however, they are performed on large labeled data which are expensive to generate. Therefore, active learning strategy is proposed in this paper to reduce the cost of annotation in COVID-19 prediction task.

Conventional active learning (i.e. pool-based active learning) has been extensively explored in the literature [18], [22]. Most of the methods operate in an iterative manner, where "the most informative sample" is chosen for labelling. Subsequently, the model is retrained with the newly labeled example. The steps are iterated alternatively until most of the examples can be classified with "reasonably high confidence" [23]. Re-training after each iteration is quite costly, especially with complex and expensive models. This is the main rationale behind batch-mode active learning methods, which select a group of informative instances simultaneously. The BMAL methods are characterized into two main groups: 1) global methods, 2) cluster-based approaches. *Global methods* try to find the most informative set of samples from the whole space directly by solving an optimization problem [20], [24]–[28]. These approaches have mathematically and empirically demonstrated a good performance, however, they do not scale well with big datasets [29]. On the other hand, *clustering-based methods*, which are highly scalable, partition either whole [30] or a fraction of (i.e. the most uncertain) unlabeled space [23], [31], [32] to reduce the probability of picking correlated queries. Once the partitions are formed, one or multiple instances are chosen to represent it.

Recently, many works have implemented a combination of deep neural networks for active learning [33]–[35]. However, to the best of our knowledge, COVID-Al [36] is the only work that explores active learning for CT scan data labeling. The authors use hybrid active learning with a 3D residual network that simultaneously considers sample diversity and predicted

loss. Despite there being many deep active learning methods that directly utilize an uncertainty-based sampling strategy, deep active learning can easily lead to insufficient diversity of batch query samples (such that relevant knowledge regarding the data distribution is not fully utilized). This in turn leads to low or even invalid Deep Learning (DL) model training performance. Thus, a feasible strategy would be to use a hybrid query strategy in a batch query, taking into account both the information volume and diversity of samples in either an explicit or implicit manner.

III. PROBLEM DEFINITION

First $X = \{x_1, x_2, \ldots, x_n\}$ denotes a dataset of n instances. Let's introduce the labeled set L and unlabeled set U, where $L \bigcup U = X$ and $L \bigcap U = \phi$. Every instance in L, x_i^L is associated with a label y_i^l, which has been revealed by a domain expert d and thus is known, whereas the labels associated with x_i^U are still unknown. The proposed approach interactively selects a batch B of samples that satisfies $B \subset U$ and $|B| = b$, where the batch size b is defined by human handling ability. Note, that all instances are of equal annotation cost. The proposed approach operates in T iterations. In each iteration, the learner will choose b instances to be labeled by the domain expert and add these labeled examples to the L to update the classifier.

IV. PROPOSED APPROACH

This section presents our proposed approach, namely, the discriminative skew-specialized sampling (DS3), which has been specifically designed to tackle the class-imbalance problem in real-world applications. The illustration of the proposed approach can be seen in Fig. 1. The framework is comprised of two main components: 1) Batch-mode imbalance learning, which predominantly focuses on finding a compromise between exploration and exploitation to effectively cover an uncertain space subject to a predefined budget and 2) Balancing approach, which addresses the unbalanced class problem for active learning.

A. BATCH-MODE IMBALANCE LEARNING

The main objective of DS3 is to develop a scalable batch-model framework for the class-imbalance problem. The success of batch mode active learning (BMAL) depends on selecting representative samples [37] as well as the batch size and total budget constraints [29]. The key question is how to find the most representative samples from both the minority and majority classes to cover the whole uncertain space given the limited budget. To achieve these goals, the DS3 learning component consists of two folds: a) Partition-based exploration and representation, and b) Skewed-specialized sampling.

1) PARTITION-BASED EXPLORATION

Dealing with massive amounts of unlabeled data, it is not feasible for a domain expert to examine every entry, and given the limited budget, it is very likely that portions of minority space are poorly represented. Therefore, it is beneficial to develop a discriminative model that can distinguish the most informative samples based on certain criteria such as ranking function. The proposed discriminative model is inspired from Guo and Schuurmans [25] work. Having access to both labeled and unlabeled samples, we built a model to maximize the expected log likelihood of the labeled data and to minimize the entropy of the missing labels on the unlabeled data:

$$\sum_{i \in L} \log P(y_i|x_i, w) + \alpha \sum_{j \in U} \sum_{y=\pm 1} P(y|x_j, w) \log P(y|x_y, w), \tag{1}$$

where w specifies the classification model, L is the labeled data, U the unlabeled instances and α is the tradeoff parameter. In order to maximize the objective function in equation 1, we construct a scoring function for a set of selected candidates S in iteration $t + 1$ according to:

$$F(S) = \sum_{i \in L^t \cup S} \log P(y_i|x_i, w^{t+1}) - \alpha \sum_{j \in U^t \setminus S} H(y|x_j, w^{t+1}) \tag{2}$$

where w^{t+1} is the parameter set for the conditional classification model trained on the new labeled set $L^{t+1} = L^t \cup S$ and $H(y|x_j, w^{t+1})$ denotes the entropy of the conditional distribution $P(y|x_j, w^{t+1})$ such that

$$H(y|x_j, w^{t+1}) = -\sum_{y=\pm 1} P(y|x_j, w^{t+1}) \log P(y|x_j, w^{t+1}) \tag{3}$$

Thus, the next strategy would be selecting a batch that has highest rank. We ranked all the samples using equation 2 and took the highest rank. We only selected highest scores (10% from the unlabeled amount). However, selecting top K data as a batch would harm the performance since many homogeneous samples would be selected due to sharing similar uncertainty scores. Thus, we used a partition-based approach for uncertain space exploration that divides the problem space into a K disjoint partition, resulting in a higher potential to explore the regions of the minority class [38]. In this work, we use K-Means and set the cluster size to 180 based on the budget we derived from a batch-selection experiment. In our experimental studies, we also examine the effect of changing the cluster size.

$$J = \sum_{j=1}^{K} \sum_{i=1}^{n} ||x_i^{(j)} - c_j||^2 \tag{4}$$

Once a cluster is formed, a representative set, which is significantly smaller from the original set, needed to be identified. J in question 4 represent the centroid of each cluster. A good representation should capture most of the information from the original set. Three samples near the centroid of the cluster is used as the represented sample, with the intuition that the central point could represent a substantial portion of the instances inside a specific partition. Furthermore, several pieces of literature mention that the clusters are represented

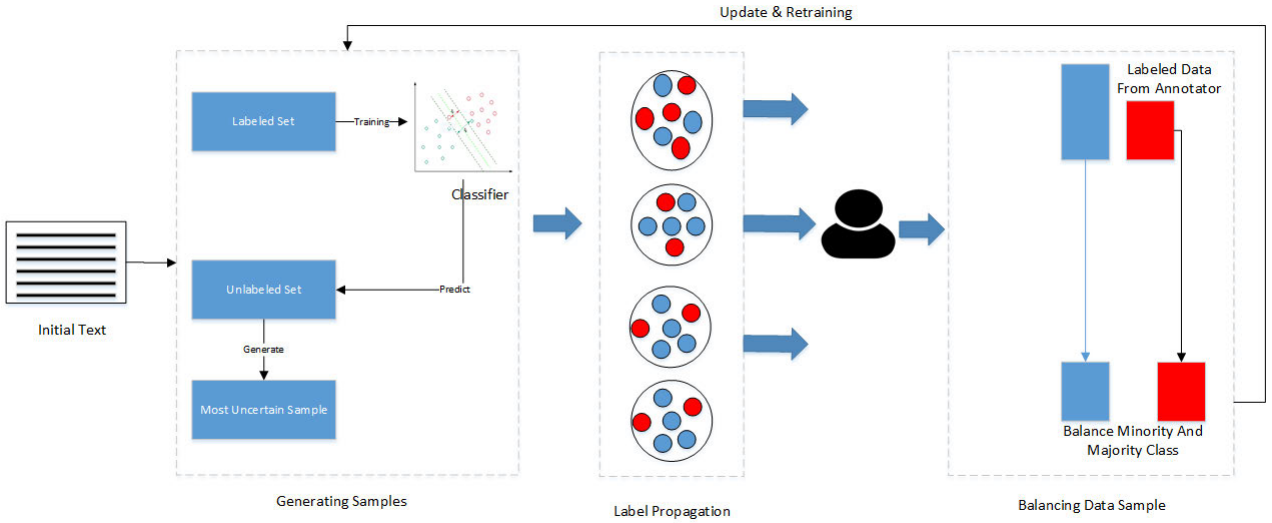


FIGURE 1. Illustration of proposed method.

by a central point [31], [32]. However, in K-Means clustering, the centroid point does not usually belong to certain samples, thus selecting three samples near the center improves the probability of selecting the most representative samples inside the cluster. Algorithm 1 summarizes our DS3 algorithm.

2) SKEWED-SPECIALIZED SAMPLING

In a highly-skewed environment where the amount of samples in the minority class is extremely low (under 10%) [19], conventional active learning approaches tend to perform poorly. It is due to this fact that even using the intelligent active learning approach, the probability of picking a minority sample is under 1% [39]. Thus, the model's performance tends to fluctuate over the training iterations. One of the classic approaches to overcoming the class imbalance is to represent the classes in a more balanced way either by oversampling the minority class, under-sampling the majority, or a blend of two approaches. Here, a simple yet effective method is proposed which maintains the original population of the minority class while under-sampling the majority class in the query set. This method keeps the model stable by selecting the best representative sample to be labeled.

B. BATCH SELECTION

Much research on batch-mode active learning picks the batch as an arbitrary number thus neglecting the real human limitation on labeling. Commonly, a batch of 20, 50, 100, 150 and 200 samples are selected as batches for labeling. However, there is lack of explanation as to how this number is selected. Therefore, this work follows studies that explore their reasoning behind selecting a specific number of batches by implementing recent studies [40], [41] that select 180 as the batch size. While Mirisaee *et al.* [40] chose 180 because it provides a good representation of the entire data, Fajri *et al.* [41] selected 180 by doing a real human labeling experiment. The work [41] shows that 180 samples is suitable for clustered text data which has a large feature space. Therefore, it is well suited for lower feature spaces such as the COVID-19 dataset presented in this paper.

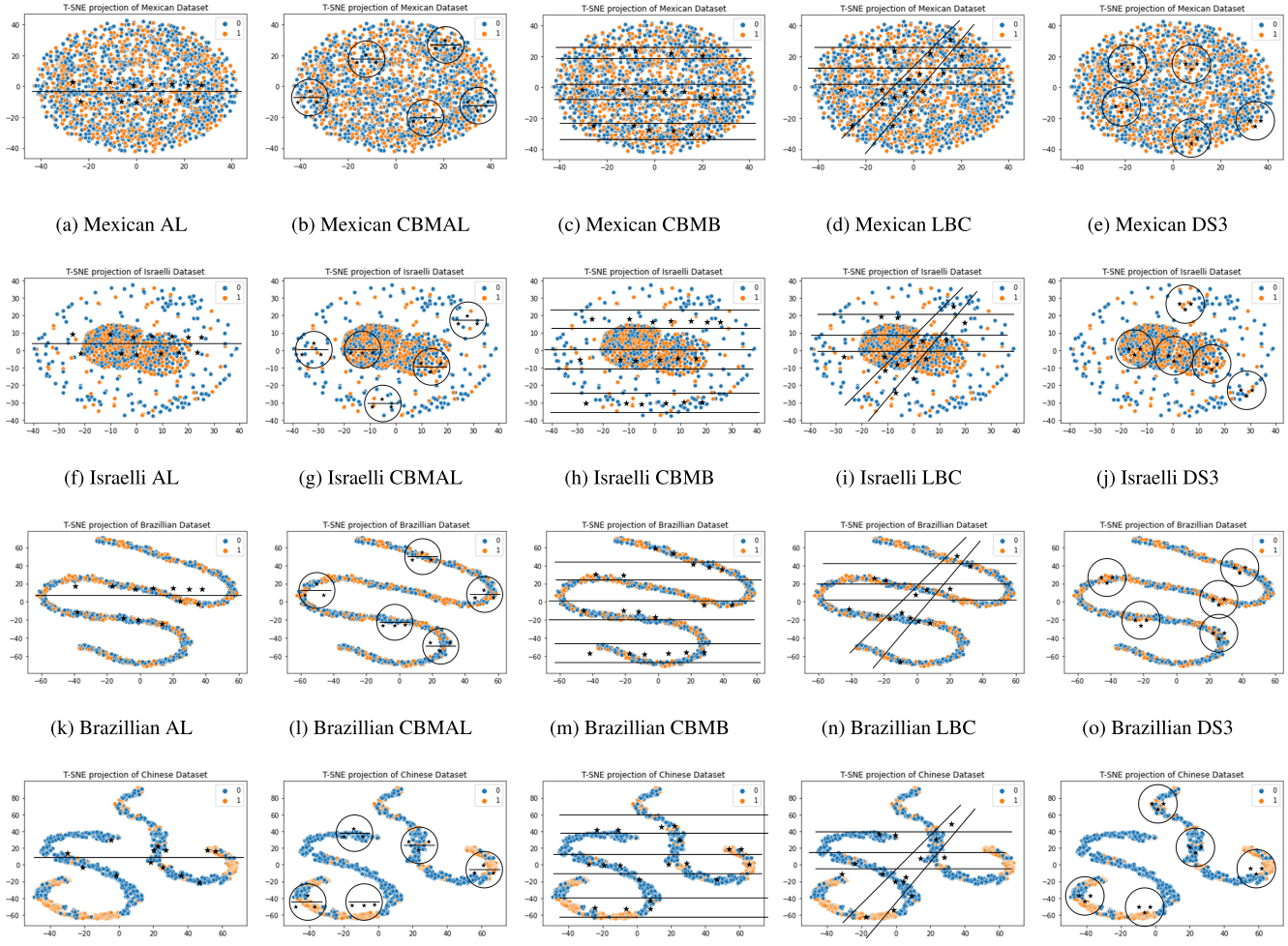
Algorithm 1: DS3 Algorithm

Input: Labeled Dataset (L)
$X = \{(x_1, y_1), \ldots, (x_n, y_n)\}$
Unlabeled Dataset $U = \{(x_1, \ldots, x_n)\}$
Labeling Budget b, Classifier $\mathcal{C}$
Output: Sample Selection $X^* = \{(x_1^*, \ldots, x_n^*) \in U\}$
initialization;
Calculate entropy $x \in C(x)$ using equation 3 ;
while $b < |B|$ **do**
  Initialize k cluster randomly;
  Set cluster prototype as cluster centroid ;
  Select representative data from cluster using equation 4 ;
  Balance the amount using random under-sampling ;
end
Sample representation selection X^* ;
Add label (x^*, y) to L and remove X^* from U ;
Update the model C_t using L ;

V. EXPERIMENTAL METHODOLOGY

A. DATASETS

Several experiments are conducted on four publicly available COVID-19 datasets from several countries. We focus primarily on COVID-19 datasets as they represent both a recent data science problem and an imbalanced set of data. Table 1 illustrates the characteristics of datasets used in this paper. As the study is designed for predicting the COVID-19 cases,



(a) Mexican AL (b) Mexican CBMAL (c) Mexican CBMB (d) Mexican LBC (e) Mexican DS3

(f) Israelli AL (g) Israelli CBMAL (h) Israelli CBMB (i) Israelli LBC (j) Israelli DS3

(k) Brazillian AL (l) Brazillian CBMAL (m) Brazillian CBMB (n) Brazillian LBC (o) Brazillian DS3

(p) Chinese AL (q) Chinese CBMAL (r) Chinese CBMB (s) Chinese LBC (t) Chinese DS3

FIGURE 2. **Comparison of five sample selection algorithms. The black star represents the selected sample.**

TABLE 1. **Datasets.**

Dataset	#Instances	#Positive	#Negative	%Positive	%Negative	#Features	Imbalance Ratio
Mexican COVID [1]	26143	101753	159670	38%	62%	24	1 : 1.5
Israeli COVID [42]	163639	69327	94312	43%	57%	10	1 : 1.3
Brazillian COVID [2]	2092	976	1116	46%	54%	12	1 : 1.14
Chinese COVID [43]	2542	912	1630	35%	65%	11	1 : 1.78

thus the dataset is designed to contain mixed features. The features range from categorical and numerical data, such as age or COVID-19 symptoms. The Table 1 shows the number of features in each dataset as well as the imbalanced ratio.

1) DATA PRE-PROCESSING AND PARAMETER TUNING

This paper follows a standard machine learning data pre-processing, including deleting the null value and performing encoding on categorical data.

B. TWO-DIMENSIONAL VISUALIZATION OF DATASETS

Further experiments compared the sample selection strategy of each model and used T-SNE [44] visualization as it able to preserve local structure of the data compared to PCA. Fig. 2 showed the illustration of the sample selection; a black star represents the selected sample, black lines represent the classification decision boundary, and circles represent a cluster. The figure showed that all samples in the datasets are non-linearly separable. Thus, nonlinear classifiers such as Random Forest, SVM, and neural network are suitable as base classifiers.

Uncertainty-based Active Learning (AL) tends to select the samples near the edge of the classification boundary. This suggests that the classifier plays an important role in sample selection; samples that are far from decision boundary have a lower probability of being picked. To increase the ability of selecting the most representative sample, CBMB [39] and

LBC [37] introduce several classifier boundaries. LBC chose several upper bounds while CBMB measured the classification cost based on uncertainty sampling. Both approaches can increase the ability of selecting the most representative sample in 'round shape' data, such as in the Mexican and Israeli datasets. However, in a spherical data shape, such as the Brazilian and Chinese datasets, a cluster-based approach, such as Certainty-based BMAL (CBMAL) and DS3, has better performance. This performance is supported by the clustering algorithm, which can locate the representative sample at the edge of the data shape.

C. LEARNING ALGORITHM & HYPERPARAMETERS SETTING

The proposed DS3 approach is a model-agnostic method; thus, any classification algorithm could be implemented. In the experiment, the Random Forest is selected as the main learning algorithm. The model extracted the entropy from class prediction probability that resulted from Random Forest as the uncertainty sampling method, shown in equation 3.

The random forest model is chosen because it is simple and shows potential performance in many machine learning problems [45], [46]. The hyperparameter of the random forest is set to have 50 numbers of trees and 4 level of depth for each tree. Several experiments with different types of classifier is performed with details in section VI-C. The other hyperparameter in the experiment is the cluster size, which is fixed to 60 by default. However, several experiments with different cluster size is also tested to evaluate the sensitivity of the proposed approach in section VI-B.

D. BASELINE METHODS

The DS3 method was compared with the most recent clustered-based BMAL approaches and the standard active learning method:

- **CBMAL (Certainty-Based BMAL)** [31]. The most ambiguous points are clustered together and the most uncertain point inside a cluster is sent for labelling.
- **AL(Active Learning)** [18]. In uncertainty-based active learning, first the model calculates the uncertainty of each sample then it presents a batch of uncertain data to be labeled.
- **LBC** [37]. As one of the most recent state-of-the-art *objective-driven* batch active learning methods, LBC uses the lower bounded certainty score of unlabeled data. Subsequently, a large similarity matrix over all unlabeled space is formed and a random greedy algorithm is employed to find a candidate batch for labeling.
- **CBMB (Cost-Bound Make-Balance** [39]. CBMB is a recent active learning approach that was implemented in unbalanced class distribution. This approach consists of two parts, Cost Bound and Make Balance. Cost Bound is used to select the candidate sample based on a cost condition (uncertainty sampling or generated sample cost) while Make Balance is used to balance the majority class samples with the amount of minority class samples. The majority sample selection is done using random strategy.

E. EVALUATION CRITERIA

In conventional classification problem, *accuracy* is a standard choice for performance evaluation. The accuracy score is straightforward and easy to implement:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{5}$$

This work implement accuracy as one of the evaluation metrics. However, it fails to reflect the performance on the skewed datasets. In such scenarios, G-Mean and F1 measures are widely used in the literature. G-Mean is the geometric mean of the accuracies of both minority and majority classes:

$$\text{G-Mean} = \sqrt{\frac{\text{TP}}{\text{TP} + \text{FN}} \times \frac{\text{TN}}{\text{TN} + \text{FP}}} \tag{6}$$

and F-1 measures output the harmonic mean of precision and recall:

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

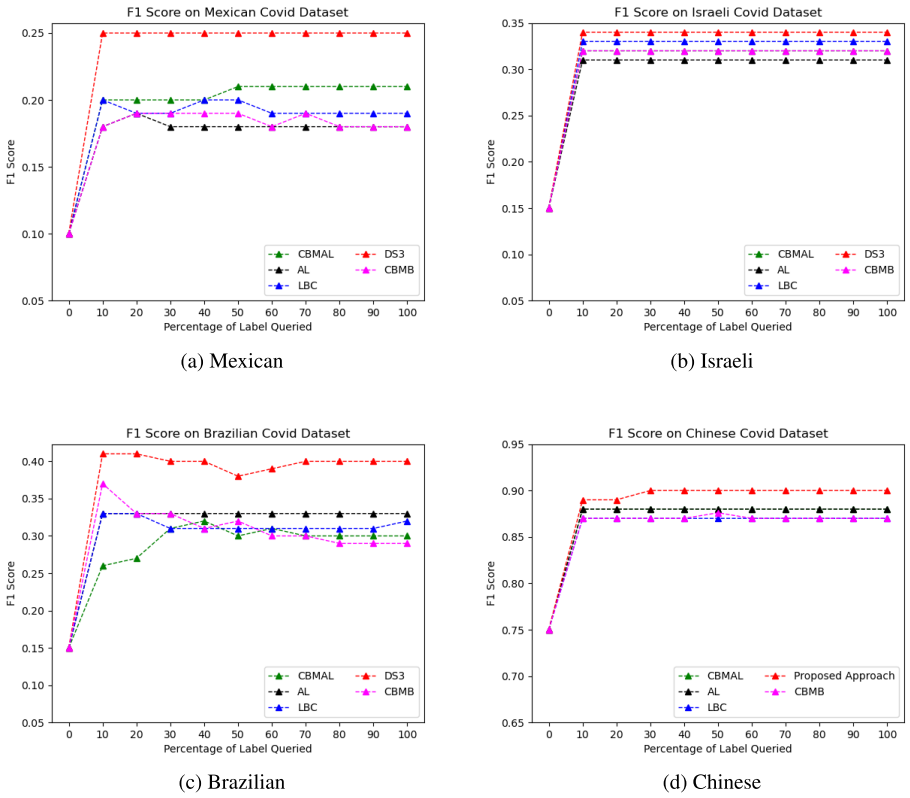
VI. RESULTS AND DISCUSSION

In this section evaluations of the performance of the proposed approach and comparisons of the model with other state-of-the-art methods is discussed. The experiments focus on the G-Mean measurement; and the investigation of why the proposed of the balancing approach outmatches other model is also presented.

A. PERFORMANCE EVALUATION

The proposed DS3 algorithm was compared with several state-of-the-art active learning models: CBMAL [31], CBMB [39] and LBC [37], and a common active learning baseline [18]. For comparison, a standard pool-based active learning strategy is implemented, dividing each dataset into 3 disjoint sets train (10%), test (20 %) and unlabeled (70 %). Fig. 3 compares the F1 score of the active learning model. The figure shows the proposed algorithm ranked first in F1 score. DS3 generally outperformed the state-of-the-art active learning models. It outperformed best when the data it was in a spherical shape, for example in Brazilian and Chinese datasets where most of the informative samples reside on the edge of the data location far away from the center. Thus, for this shape, a clustering-based active learning approach performed well in selecting the most informative sample. In some cases, for example in Mexican and Israeli datasets, DS3 performed equally well with the comparison method, with a few advantages. For example, DS3 reached a 0.25 and 0.34 F1 score on the Mexican and Israeli datasets respectively, having a 0.10 difference compared to the baselines.

For evaluation purposes, the performance comparison of each model in is presented on Table 2. Based on the table, the proposed approach shows a high performance compared



(a) Mexican (b) Israeli

(c) Brazilian (d) Chinese

FIGURE 3. Comparison of the performance of DS3 with baseline methods on multiple datasets (F1 against number of queries).

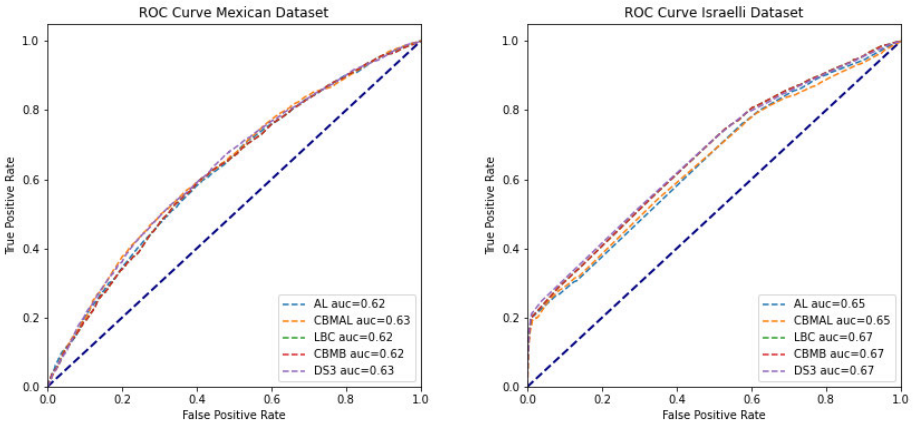
to other state-of-the-art models. This performance was highlighted in three different evaluation metrics. For example, in the Chinese COVID-19 dataset, DS3 wins in all evaluation metrics, however, in the Israeli dataset it only succeeds in 2 categories (i.e., F1 Score and accuracy). Despite DS3 losing the G-Mean score compared to the Israeli dataset, the difference is only 0.01. To further highlight the performance of each method, a ROC curve is discussed. A similar Random Forest classifier is selected for the experiment. Fig. 4 shows the result of each dataset with respect to the ROC curve. Almost all models perform equally well in the Mexican, Israeli, and Chinese datasets. The performance differences on these datasets are only marginal. In the Brazilian dataset, DS3-labeled data showed a higher ROC Curve with 0.71, which is 0.11 points above LBC and CBMB.

B. EXPERIMENTS USING DIFFERENT CLUSTER SIZES

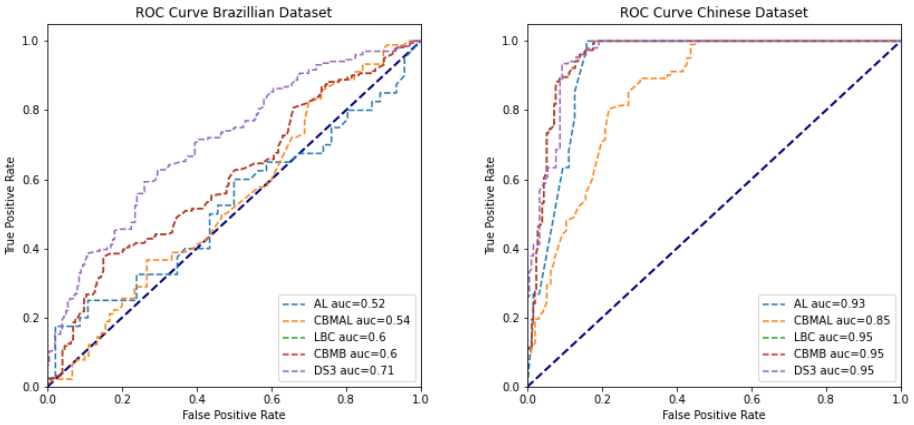
A further experimentation of the robustness of the DS3 approach using different cluster sizes is conducted. The main objective of this experiment was to explore whether cluster size has a significant impact on the approach. In CBMAL approaches, such as DS3, the cluster size selection will influence the representation of the data. Thus, the choice of cluster size should maximize the creation a homogeneous cluster, leading to the ease of selecting representative data and contributing to the better model performance. Table 3 illustrates DS3 performance across different cluster sizes. In particular, it shows that the performance of the DS3 model increases with increased cluster size. For example, in the Mexican COVID-19 dataset, DS3 with cluster size of 100 has a G-Mean of 0.59 and an F1 score of 0.20. This number rises when the cluster size is extended to 300. Each model gains performance, reaching a G-Mean of 0.65 and an F1 Score of 0.37. However, in examining the result, one can infer that the accuracy of DS3 in the Israeli dataset behaves opposite to the general result. This could be influenced by the characteristic of the Israeli dataset, which is the largest dataset wherein 180 samples selected from the cluster could not represent the dataset well.

C. EXPERIMENTS USING DIFFERENT CLASSIFIERS

The DS3 underlying classifier was compared with other well-known tree-based algorithms such as AdaBoost, CatBoost, XGBoost, and LightGBM. The Random Forest was chosen as the underlying classifier of our main algorithm. Previously, Support Vector Machine was a popular classifier for active learning [47], [48]; however, many recent works prefer Random Forest as the base classifier since it works well for BMAL in unbalanced class distribution [49]–[51]. Since most of the datasets have an imbalanced class ratio, Random Forest was chosen as base classifier. The results in Table 4 compare the performance of each classifier with regards



(a) Mexican (b) Israeli



(c) Brazilian (d) Chinese

FIGURE 4. ROC curve of DS3 with baseline methods on multiple datasets.

TABLE 2. Overall model performance (Average over 10 runs | standard deviation).

Methods	Mexican COVID			Israelli COVID			Brazillian COVID			Chinese COVID		
	G-Mean	F1 Score	Accuracy	G-Mean	F1 Score	Accuracy	G-Mean	F1 Score	Accuracy	G-Mean	F1 Score	Accuracy
CBMAL	0.59 | 0.004	0.20 | 0.016	**0.62 | 0.011**	0.74 | 0.016	0.32 | 0.009	**0.65 | 0.003**	0.54 | 0.042	0.30 | 0.065	0.53 | 0.013	**0.89 | 0.014**	0.88 | 0.014	0.88 | 0.014
AL	0.59 | 0.004	0.19 | 0.018	0.61 | 0.004	0.73 | 0.047	0.31 | 0.047	0.63 | 0.020	0.55 | 0.023	0.33 | 0.079	0.54 | 0.020	**0.89 | 0.010**	0.88 | 0.011	0.87 | 0.009
LBC	0.59 | 0.004	0.19 | 0.015	0.61 | 0.009	**0.75 | 0.004**	0.33 | 0.008	0.64 | 0.0097	0.56 | 0.022	0.31 | 0.069	0.55 | 0.016	0.88 | 0.019	0.87 | 0.021	0.86 | 0.022
CBMB	**0.60 | 0.006**	0.18 | 0.017	0.61 | 0.003	0.72 | 0.04	0.32 | 0.009	0.64 | 0.003	0.53 | 0.035	0.26 | 0.086	0.54 | 0.019	0.88 | 0.020	0.87 | 0.022	0.87 | 0.024
DS3	0.59 | 0.004	**0.35 | 0.014**	**0.62 | 0.004**	0.74 | 0.008	**0.34 | 0.012**	**0.65 | 0.002**	**0.57 | 0.024**	**0.34 | 0.057**	**0.56 | 0.014**	**0.89 | 0.008**	**0.89 | 0.010**	**0.90 | 0.008**

TABLE 3. Model performance using different cluster sizes.

Cluster Size	Dataset																							
	Mexican COVID						Israeli COVID						Brazillian COVID						Chinese COVID					
	G-Mean		F1 Score		Accuracy		G-Mean		F1 Score		Accuracy		G-Mean		F1 Score		Accuracy		G-Mean		F1 Score		Accuracy	
	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD
100	0.595	0.0048	0.2	0.01	0.62	0	0.75	0.004	0.32	0.004	0.64	0.008	0.53	0.03	0.45	0.03	0.53	0.03	0.85	0.02	0.841	0.03	0.86	0.02
120	0.596	0.005	0.2	0.01	0.62	0	0.752	0.004	0.328	0.006	**0.647**	0.006	0.538	0.01	0.443	0.03	0.54	0.01	0.843	0.01	0.836	0.02	0.851	0.01
140	0.599	0.005	0.2	0.01	0.622	0.04	0.756	0.0004	0.328	0.007	**0.647**	0.006	0.54	0.01	0.456	0.05	0.54	0.01	0.849	0.02	0.841	0.02	0.869	0.01
160	0.596	0.0048	0.2	0.01	0.622	0.04	0.76	0.0018	0.33	0.005	0.645	0.008	0.551	0.02	0.45	0.03	0.53	0.01	0.843	0.02	0.828	0.02	0.848	0.01
180	0.601	0.01	0.25	0.002	0.621	0.003	0.762	0.0018	0.332	0.003	0.643	0.006	0.558	0.02	0.49	0.04	0.54	0.03	0.854	0.02	0.846	0.02	0.863	0.02
200	0.61	0.013	0.26	0.01	0.621	0.003	0.77	0.005	0.336	0.005	0.641	0.009	0.561	0.02	0.5	0.05	0.537	0.03	0.858	0.01	0.845	0.01	0.863	0.02
220	0.616	0.02	0.32	0.01	0.621	0.003	0.772	0.001	0.34	0.001	0.643	0.008	0.564	0.04	0.509	0.03	0.51	0.03	0.839	0.02	0.823	0.02	0.864	0.01
240	0.625	0.02	0.33	0.02	**0.626**	0.004	0.776	0.008	0.34	0.009	0.641	0.009	0.571	0.03	0.511	0.04	**0.557**	0.02	0.846	0.01	0.838	0.01	0.85	0.01
260	0.63	0.03	0.36	0.008	0.624	0.004	0.78	0.05	0.343	0.009	0.64	0.01	0.573	0.04	0.52	0.05	0.53	0.02	0.855	0.01	0.843	0.01	0.86	0.01
280	0.64	0.04	0.36	0.001	0.62	0	0.781	0.008	0.346	0.008	0.63	0.01	0.583	0.05	0.523	0.05	0.526	0.02	0.847	0.01	0.832	0.02	0.854	0.02
300	**0.65**	0.05	**0.37**	0.008	0.621	0.003	**0.786**	0.007	**0.35**	0.006	0.63	0.01	**0.597**	0.05	**0.538**	0.04	0.54	0.03	**0.862**	0.02	**0.857**	0.02	**0.876**	0.01

to the G-Mean, F1 Score, and Accuracy score. The results show that DS3 with Random Forest performs slightly better compared to other base classifiers. For example, Random Forest has a better F1 score in the Mexican and Chinese COVID-19 datasets, and DS3 with Random Forest reaches an F1 score of 0.37 and 0.89 in the Mexican and Chinese

TABLE 4. Model performance using different classifiers.

Main Algorithm	Mexican COVID			Israeli COVID			Brazilian COVID			Chinese COVID		
	G-Mean (AVG | STD)	F1 Score (AVG |STD)	Accuracy (AVG | STD)	G-Mean (AVG | STD)	F1 Score (AVG | STD)	Accuracy (AVG | STD)	G-Mean (AVG | STD)	F1 Score (AVG | STD)	Accuracy (AVG | STD)	G-Mean (AVG | STD)	F1 Score (AVG | STD)	Accuracy (AVG | STD)
AdaBoost	**0.59 | 0.004**	0.34 | 0.0127	**0.62 | 0.005**	0.73 | 0.005	0.33 | 0.004	**0.64 | 0.003**	0.52 | 0.02	0.50 | 0.03	0.53 | 0.02	0.89 | 0.01	0.87 | 0.01	0.90 | 0.009
XGBoost	0.58 | 0.0049	0.35 | 0.014	0.61 | 0.001	0.67 | 0.007	**0.37 | 0.006**	0.63 | 0.005	0.52 | 0.002	0.52 | 0.02	0.52 | 0.02	0.89 | 0.01	0.88 | 0.03	**0.91 | 0.01**
CatBoost	**0.59 | 0.0028**	0.31 | 0.032	**0.62 | 0.003**	**0.75 | 0.004**	0.32 | 0.004	**0.64 | 0.009**	0.53 | 0.02	0.47 | 0.05	0.53 | 0.02	0.85 | 0.01	0.83 | 0.01	0.86 | 0.01
LightGBM	**0.59 | 0.003**	0.36 | 0.014	**0.62 | 0.003**	0.73 | 0.008	0.34 | 0.006	**0.64 | 0.003**	0.52 | 0.02	**0.53 | 0.02**	0.52 | 0.02	**0.90 | 0.01**	0.88 | 0.01	**0.91 | 0.01**
Random Forest	**0.59 | 0.0046**	**0.37 | 0.005**	**0.62 | 0.004**	0.74 | 0.008	0.34 | 0.01	**0.64 | 0.004**	**0.57 | 0.02**	0.52 | 0.02	**0.56 | 0.01**	0.89 | 0.008	**0.89 | 0.01**	0.90 | 0.008

TABLE 5. Experiments with different amounts of initial training data.

Amount of Initial Training Set	Methods									
	AL		CBMAL		CBMB		DS3		LBC	
	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD
10K	**25.18**	0.58	267.17	10.02	39.49	10.01	531.26	17.56	*704.61*	24.82
20K	**28.28**	0.89	258.03	7.38	42.19	1.73	536.15	8.09	*610.25*	7.98
40K	**35.48**	1.04	232.58	8.24	47.21	5.23	498.03	3.52	*576.23*	77.28
60K	**40.14**	0.38	218.82	17.80	56.70	5.59	452.14	17.77	*511.38*	26.85
80K	**49.67**	1.74	195.27	24.68	73.20	2.18	452.58	6.03	*481.25*	23.16
100K	**55.64**	2.09	187.34	40.44	87.65	2.33	436.15	16.34	*451.14*	9.45
120K	**65.78**	2.88	160.02	43.13	97.01	5.07	401.12	5.04	*390.88*	16.59
140K	**81.48**	0.52	171.23	5.24	109.85	8.25	360.18	15.76	*456.06*	56.61
160K	**88.38**	2.32	163.54	5.76	128.59	4.63	362.34	9.86	*406.35*	59.22
180K	**98.94**	2.55	151.99	2.68	137.10	10.13	348.04	17.23	*399.44*	21.11

TABLE 6. p-values of wilcoxon test using the random forest classifier vs other classifiers.

Random Forest Vs	Dataset			
	Mexican	Israeli	Brazilian	Chinese
AdaBoost	0.011	0.011	0.012	0.012
XGBoost	0.011	0.012	0.261	0.12
CatBoost	0.012	0.010	0.012	0.012
LightBM	0.011	**0.074**	**0.51**	0.012

datasets respectively. In other datasets, Random Forest performs equally well compared to other approaches. However, the performance of DS3 Random Forest is lower than XGBoost and LightBM in both the Israeli and Brazilian datasets; nonetheless the differences between Random Forest and these classifiers is only 0.01 and thus does not illustrate the real performance of DS3 with Random Forest.

Further examination of the performance of each classifier is measured statistically. First a test of the normality assumption is conducted by performing the Kolmogorov-Smirnov test, and the F1 score is used as the base score for statistic evaluation. The results showed that the test sample failed the normality test, thus non-parametric tests such as the Wilcoxon test, are more appropriate for evaluating the performance of our model statistically. We used the Wilcoxon signed-rank test, and the results are presented in Table 6.

The Wilxocon test shows that in almost all datasets Random Forest is significantly better than AdaBoost, XGBoost, CatBost, and LightBM. However, there are no significant differences between Random Forest and LightBM in the Israeli and Brazilian datasets.

D. EXPERIMENTS USING DIFFERENT INITIAL AMOUNTS OF TRAINING DATA

A final experiment was conducted to evaluate how our DS3 framework behaves with different amount of training data. The initial training data was set to a default of 10% of the dataset; however, it is interesting to evaluate the time performance with increase amounts of initial training set data. Highlights of the execution time of each models with respect to the various amount of training data is presented in Table 5. The value in bold presents the lowest execution time while the highest running time is highlighted in italics. The experiment only use the Mexican COVID-19 dataset since it is the largest dataset presented in this paper. Table 5 also illustrates that traditional AL has the lowest execution time across all the training sets. This is intuitive since AL does not have any extra computational costs other than calculating the entropy. The LBC model shows the highest time performance since its characteristic is the opposite of AL with a lot of extra computational costs. The other three methods CBMAL, CBMB, and DS3 show an extra computational time compared to AL. CBMB has lower execution time compared to the CBMAL and DS3, while DS3 shows the highest. However, considering the results, DS3 has a better F1 score in the Mexican dataset compared to the other approaches. DS3 has an F1 score of 0.35, which almost doubled the performance compared to the rest of the models used on the Mexican dataset.

VII. DISCUSSION

This research shows that a discriminative-based approach works best for batch-mode active learning in imbalanced data scenarios. There are several potential explanations for the results. The first potential explanation is that the ability of DS3 to select the most representative data; since data that belong to the minority class inside the cluster leads to better sample selection compared to other models. The second potential explanation is that our balancing mechanism leads to more stable performance.

It is envisaged that our framework will have a positive impact on the community as our model could be used as a solution to reduce COVID-19 data annotation cost. Secondly,

with reduced cost, the deep learning model can be trained for automatic COVID-19 detection efficiently. Finally, our framework could be transferred to other domains that focus on reducing the cost of annotation in both balanced and imbalanced datasets.

Although showing a better performance compared to other baselines, the purposed approach is no silver bullet. Thus, there is room for improvement in the proposed framework. For example, in the DS3 balancing approach, the data selection is at random. In future works, it would be interesting to see how other sampling methods would behave when combined with partition-based models.

VIII. CONCLUSION

This paper proposes a discriminative batch-mode active learning framework, called DS3, for the diagnosis of COVID-19. The framework can greatly reduce the cost of manual labeling for training models and can further relieve the burden of the healthcare system in the case of a fast-spreading pandemic. The proposed framework can boost the performance of any machine learning model by simultaneously considering diversity and representativeness of the data samples that also fit the imbalanced data distribution. To verify the effectiveness of DS3, extensive experiments have been conducted on various real-world COVID-19 datasets. The experimental and statistical significance test results demonstrate that the DS3 outperforms the baselines of state-of-the-art batch-mode active learning methods.

ACKNOWLEDGMENT

(Amril Nazir and Ricky Maulana contributed equally to this work.)

REFERENCES

[1] A. Kucharski, T. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, and R. Eggo, "Early dynamics of transmission and control of COVID-19: A mathematical modelling study," *Lancet Infectious Diseases*, vol. 20, pp. 553–558, May 2020.

[2] C. E. Lopez, M. Vasu, and C. Gallemore, "Understanding the perception of COVID-19 policies by mining a multilanguage Twitter dataset," 2020, *arXiv:2003.10359*.

[3] L. L. Wang, K. Lo, Y. Chandrasekhar, R. Reas, and J. Yang, "CORD-19: The COVID-19 open research dataset," 2020, *arXiv:2004.10706*.

[4] A. Imran, I. Posokhova, H. N. Qureshi, U. Masood, S. Riaz, K. Ali, C. N. John, and M. Nabeel, "AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app," *Inf. Med. Unlocked*, vol. 20, Jan. 2020, Art. no. 100378.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[6] G. G. Patil and R. K. Banyal, "Techniques of deep learning for image recognition," in *Proc. IEEE 5th Int. Conf. Converg. Technol. (I2CT)*, Mar. 2019, pp. 1–5.

[7] C. Chen, O. Li, A. Barnett, J. K. Su, and C. Rudin, "This looks like that: Deep learning for interpretable image recognition," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 1–12.

[8] N. H. Phong and B. Ribeiro, "Offline and online deep learning for image recognition," in *Proc. 4th Exp. Int. Conf.*, Jun. 2017, pp. 171–175.

[9] Y. Lin, Y. Meng, X. Sun, Q. Han, K. Kuang, J. Li, and F. Wu, "BertGCN: Transductive text classification by combining GCN and BERT," in *Proc. Findings Assoc. Comput. Linguistics*, 2021, pp. 1456–1462.

[10] Z. Yang, Z. Dai, Y. Yang, J. G. Carbonell, R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized autoregressive pretraining for language understanding," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 1–4.

[11] L. Yao, C. Mao, and Y. Luo, "Graph convolutional networks for text classification," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 7370–7377.

[12] L. Deng and J. C. Platt, "Ensemble deep learning for speech recognition," in *Proc. INTERSPEECH*, 2014, pp. 1–6.

[13] J. Dai, Y. Zhang, J. Hou, X. E. Wang, L. Tan, and J. Jiang, "Sparse wavelet decomposition and filter banks with CNN deep learning for speech recognition," in *Proc. IEEE Int. Conf. Electron. Inf. Technol. (EIT)*, May 2019, pp. 098–103.

[14] K. Jermsittiparsert, A. Abdurrahman, P. Siriattakul, L. A. Sundeeva, W. Hashim, R. Rahim, and A. Maseleno, "Pattern recognition and features selection for speech emotion recognition model using deep learning," *Int. J. Speech Technol.*, vol. 23, no. 4, pp. 799–806, Dec. 2020.

[15] A. S. Al-Waisy, S. Al-Fahdawi, M. A. Mohammed, K. H. Abdulkareem, S. A. Mostafa, M. S. Maashi, M. Arif, and B. Garcia-Zapirain, "COVID-CheXNet: hybrid deep learning framework for identifying COVID-19 virus in chest X-rays images," *Soft Comput.*, vol. 2020, pp. 1–16, Nov. 2020, doi: 10.1007/s00500-020-05424-3.

[16] A. S. Al-Waisy, M. A. Mohammed, S. Al-Fahdawi, M. S. Maashi, B. Garcia-Zapirain, K. H. Abdulkareem, S. A. Mostafa, N. M. Kumar, and D. N. Le, "COVID-DeepNet: Hybrid multimodal deep learning system for improving COVID-19 pneumonia detection in chest X-ray images," *Comput., Mater. Continua*, vol. 67, no. 2, 2021, pp. 2409–2429, 2021.

[17] M. A. Mohammed, K. H. Abdulkareem, A. S. Al-Waisy, S. A. Mostafa, S. Al-Fahdawi, A. M. Dinar, W. Alhakami, A. Baz, M. N. Al-Mhiqani, H. Alhakami, N. Arbaiy, M. S. Maashi, A. A. Mutlag, B. García-Zapirain, and I. de la Torre Díez, "Benchmarking methodology for selection of optimal COVID-19 diagnostic model based on entropy and topsis methods," *IEEE Access*, vol. 8, pp. 99115–99131, 2020.

[18] B. Settles, "Active learning literature survey," Comput. Sci., Univ. Wisconsin–Madison, Madison, WI, USA, Tech. Rep. 1648, 2009.

[19] J. Attenberg and F. J. Provost, "Why label when you can search?: Alternatives to active learning for applying human resources to build classification models under extreme class imbalance," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2010, pp. 423–432. [Online]. Available: https://doi.org/10.1145/1835804.1835859

[20] S. Chakraborty, V. Balasubramanian, and S. Panchanathan, "Adaptive batch mode active learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 8, pp. 1747–1760, Aug. 2015.

[21] M. A. Mohammed, K. H. Abdulkareem, B. Garcia-Zapirain, S. A. Mostafa, M. S. Maashi, A. S. Al-Waisy, M. A. Subhi, A. A. Mutlag, and D.-N. Le, "A comprehensive investigation of machine learning feature extraction and classification methods for automated diagnosis of COVID-19 based on X-ray images," *Comput., Mater. Continua*, vol. 66, no. 3, pp. 3289–3310, 2021.

[22] H. Yu, X. Yang, S. Zheng, and C. Sun, "Active learning from imbalanced data: A solution of online weighted extreme learning machine," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 4, pp. 1088–1103, Apr. 2019.

[23] X. Xia, P. Protopapas, and F. Doshi-Velez, "Cost-sensitive batch mode active learning: Designing astronomical observation by optimizing telescope time and telescope choice," in *Proc. SIAM Int. Conf. Data Mining*, Miami, FL, USA, Jun. 2016, pp. 477–485.

[24] G. Schohn and D. Cohn, "Less is more: Active learning with support vector machines," in *Proc. 17th Int. Conf. Mach. Learn.*, 2000, pp. 839–846.

[25] Y. Guo and D. Schuurmans, "Discriminative batch mode active learning," in *Proc. 20th Int. Conf. Neural Inf. Process. Syst.*, 2007, pp. 593–600.

[26] S. C. H. Hoi, R. Jin, and M. R. Lyu, "Batch mode active learning with applications to text categorization and image retrieval," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1233–1248, Sep. 2009.

[27] S. Chakraborty, V. Balasubramanian, and S. Panchanathan, "Dynamic batch mode active learning," in *Proc. CVPR*, Jun. 2011, pp. 2649–2656.

[28] M. Wang, Y.-Y. Zhang, and F. Min, "Active learning through multi-standard optimization," *IEEE Access*, vol. 7, pp. 56772–56784, 2019.

[29] I. Lourentzou, D. Gruhl, and S. Welch, "Exploring the efficiency of batch active learning for human-in-the-loop relation extraction," in *Proc. Companion Web Conf. Web Conf.*, 2018, pp. 1131–1138, doi: 10.1145/3184558.3191546.

[30] A. Singla and S. Patra, "A fast partition-based batch-mode active learning technique using SVM classifier," *Soft Comput.*, vol. 22, no. 14, pp. 4627–4637, Jul. 2018.

[31] S. Patra and L. Bruzzone, "A cluster-assumption based batch mode active learning technique," *Pattern Recognit. Lett.*, vol. 33, no. 9, pp. 1042–1048, Jul. 2012.
[32] B. Demir, C. Persello, and L. Bruzzone, "Batch-mode active-learning methods for the interactive classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 1014–1031, Mar. 2011.
[33] N. Asghar, P. Poupart, X. Jiang, and H. Li, "Deep active learning for dialogue generation," in *Proc. 6th Joint Conf. Lexical Comput. Semantics*, 2017, pp. 78–83.
[34] T. He, X. Jin, G. Ding, L. Yi, and C. Yan, "Towards better uncertainty sampling: Active learning with multiple views for deep convolutional neural network," in *Proc. IEEE Int. Conf. Multimedia Expo. (ICME)*, Jul. 2019, pp. 1360–1365.
[35] H. Ranganathan, H. Venkateswara, S. Chakraborty, and S. Panchanathan, "Deep active learning for image classification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3934–3938.
[36] X. Wu, C. Chen, M. Zhong, J. Wang, and J. Shi, "COVID-AL: The diagnosis of COVID-19 with deep active learning," *Med. Image Anal.*, vol. 68, Dec. 2020, Art. no. 101913.
[37] H. Wang, R. Zhou, and Y.-D. Shen, "Bounding uncertainty for active batch selection," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 5240–5247, doi: 10.1609/aaai.v33i01.33015240.
[38] S. Ertekin, J. Huang, and C. L. Giles, "Active learning for class imbalance problem," in *Proc. 30th Annu. Int. Conf. Res. Develop. Inf. Retr.*, Amsterdam, The Netherlands, Jul. 2007, pp. 823–824.
[39] C. H. Lin, Mausam, and D. S. Weld, "Active learning with unbalanced classes and example-generation queries," in *Proc. 6th AAAI Conf. Hum. Comput. Crowdsourcing*, 2018, pp. 98–107.
[40] S. H. Mirisaee, A. Douzal, and A. Termier, "Selecting representative instances from datasets," in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2015, pp. 1–10.
[41] R. M. Fajri, S. Khoshrou, R. Peharz, and M. Pechenizkiy, "PS3: Partition-based skew-specialized sampling for batch mode active learning in imbalanced text data," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, 2021, pp. 68–84.
[42] Y. Zoabi, S. Deri-Rozov, and N. Shomron, "Machine learning-based prediction of COVID-19 diagnosis based on symptoms," *NPJ Digit. Med.*, vol. 4, no. 1, pp. 1–5, Dec. 2021.
[43] W. Ning, S. Lei, J. Yang, Y. Cao, P. Jiang, Q. Yang, J. Zhang, X. Wang, F. Chen, Z. Geng, L. Xiong, H. Zhou, Y. Guo, Y. Zeng, H. Shi, L. Wang, Y. Xue, and Z. Wang, "Open resource of clinical data from patients with pneumonia for the prediction of COVID-19 outcomes via deep learning," *Nature Biomed. Eng.*, vol. 4, no. 12, pp. 1197–1207, Dec. 2020.
[44] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[45] C. Iwendi, A. K. Bashir, A. Peshkar, R. Sujatha, J. M. Chatterjee, S. Pasupuleti, R. Mishra, S. Pillai, and O. Jo, "COVID-19 patient health prediction using boosted random forest algorithm," *Frontiers Public Health*, vol. 8, p. 357, Jul. 2020.
[46] M. Belgiu and L. Drăgut, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31, Apr. 2016.
[47] S. C. H. Hoi, R. Jin, and M. R. Lyu, "Large-scale text categorization by batch mode active learning," in *Proc. 15th Int. Conf. World Wide Web*, 2006, pp. 633–642, doi: 10.1145/1135777.1135870.
[48] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *J. Mach. Learn. Res.*, vol. 2, pp. 45–66, Nov. 2001.
[49] H. T. Nguyen, J. Yadegar, B. Kong, and H. Wei, "Efficient batch-mode active learning of random forest," in *Proc. IEEE Stat. Signal Process. Workshop (SSP)*, Dec. 2012, pp. 596–599.
[50] Y. Chen and S. Mani, "Active learning for unbalanced data in the challenge with multiple models and biasing," in *Proc. Mach. Learn. Res.*, vol. 16, 2011, pp. 113–126.
[51] Y. Gu, D. Zydek, and Z. Jin, "Active learning based on random forest and its application to terrain classification," in *Progress in Systems Engineering*. Cham, Switzerland: Springer, 2014, pp. 273–278.

AMRIL NAZIR is currently the Chief Architect at Codecompass Llp and an Associate Professor with the Department of Information Systems, College of Technological Innovation, Zayed University. He was formerly a Senior Research Scientist at the Malaysian Research and Development Institute for nine years. His research interests include artificial intelligence (AI), machine learning, data science, and big data.

RICKY MAULANA FAJRI received the M.Sc. degree in computer science from the University of Technology Sydney, Australia, in 2011. He is currently pursuing the Ph.D. degree with the Data Mining Group, Eindhoven University of Technology, The Netherlands. Since 2015, he has been working as a Lecturer and a Researcher with the Department of Information Technology, Faculty of Computer Science, University of Indo Global Mandiri. Palembang, Indonesia. His research interests include machine learning, active learning, and fairness for active learning.