IEEE ComSoc IEEE Communications Society | IEEE COMPUTER SOCIETY | IEEE Signal Processing Society | VTS | IEEE Transactions on Machine Learning in Communications and Networking

Received 31 July 2023; revised 10 December 2023; accepted 23 February 2024.
Date of publication 6 March 2024; date of current version 19 March 2024.

The associate editor coordinating the review of this article and approving it for publication was J. J. P. C. Rodrigues.

Digital Object Identifier 10.1109/TMLCN.2024.3374253

An IoMT-Based Incremental Learning Framework With a Novel Feature Selection Algorithm for Intelligent Diagnosis in Smart Healthcare

SIVA SAI[1], KARTIKEY SINGH BHANDARI[2], ADITYA NAWAL[3], VINAY CHAMOLA[1,4] (Senior Member, IEEE), AND BIPLAB SIKDAR[5] (Senior Member, IEEE)

[1]Department of Electrical and Electronics Engineering, Birla Institute of Technology and Science, Pilani (BITS Pilani), Pilani Campus, Pilani 333031, India
[2]Department of Computer Science and Information Systems, Birla Institute of Technology and Science, Pilani (BITS Pilani), Pilani Campus, Pilani 333031, India
[3]Microsoft Campus, Gachibowli, Hyderabad, Telangana 500032, India
[4]Anuradha and Prashanth Palakurthi Centre for Artificial Intelligence Research (APPCAIR), Birla Institute of Technology and Science, Pilani (BITS Pilani), Pilani Campus, Pilani 333031, India
[5]Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583

CORRESPONDING AUTHOR: V. CHAMOLA (vinay.chamola@pilani.bits-pilani.ac.in)

The work of Vinay Chamola was supported by the Chanakya Fellowship Program of Technology Innovation Hub (TIH) Foundation for Internet of Things (IoT) and Internet of Everything (IoE) (TIH-IoT) under Project CFP/2022/027.

ABSTRACT Several recent research papers in the Internet of Medical Things (IoMT) domain employ machine learning techniques to detect data patterns and trends, identify anomalies, predict and prevent adverse events, and develop personalized patient treatment plans. Despite the potential of machine learning techniques in IoMT to revolutionize healthcare, several challenges remain.The conventional machine learning models in the IoMT domain are static in that they were trained on some datasets and are being used for real-time inferencing data. This approach does not consider the patient's recent health-related data. In the conventional machine learning models paradigm, the models must be re-trained again, even to incorporate a few sets of additional samples. Also, since the training of the conventional machine learning models generally happens on cloud platforms, there are also risks to security and privacy. Addressing these several issues, we propose an edge-based incremental learning framework with a novel feature selection algorithm for intelligent diagnosis of patients. The approach aims to improve the accuracy and efficiency of medical diagnosis by continuously learning from new patient data and adapting to patient conditions over time, along with reducing privacy and security issues. Addressing the issue of excessive features, which might increase the computational burden on incremental models, we propose a novel feature selection algorithm based on bijective soft sets, Shannon entropy, and TOPSIS(Technique for Order Preference by Similarity to Ideal Solution). We propose two incremental algorithms inspired by Aggregated Mondrian Forests and Half-Space Trees for classification and anomaly detection. The proposed model for classification gives an accuracy of 87.63%, which is better by 13.61% than the best-performing batch learning-based model. Similarly, the proposed model for anomaly detection gives an accuracy of 97.22%, which is better by 1.76% than the best-performing batch-based model. The proposed incremental algorithms for classification and anomaly detection are 9X and 16X faster than their corresponding best-performing batch learning-based models.

INDEX TERMS Incremental learning, TOPSIS, Shannon entropy, bijective soft-sets, IoMT.

I. INTRODUCTION

IN RECENT years, the Internet of Things (IoT) in healthcare has increased exponentially. The IoT in healthcare is called the Internet of Medical Things (IoMT). The IoMT models comprise a network of medical devices and applications connected through the internet. IoMT enables the

provision of real-time patient data to healthcare professionals, which facilitates remote monitoring. The availability of real-time data helps in making informed decisions quickly. Thus, IoMT is revolutionizing the healthcare domain by improving patient-doctor communication, reducing healthcare expenses, and increasing the efficiency of medical processes [1].

According to MarketsandMarkets report [2], the worldwide market for medical devices employing the Internet of Things (IoT) is predicted to grow to USD 94.2 billion by 2026, which was at USD 26.5 billion in 2021, at a compound annual growth rate of 28.9%. Reducing healthcare costs and actively monitoring patients is causing this growth. The global IOT healthcare market [3] is expected to grow with a CAGR of 17.8% from 127 billion USD in 2023 to 289 billion USD in 2028. Advancements in medical devices and healthcare infrastructure and digital transformation across the industry are the primary drivers for this.

In IoMT, various machine-learning algorithms [4], [5] are used to analyze and interpret large amounts of data from medical devices like wearables, sensors, and monitoring systems. These algorithms make use of statistical models which help in detecting patterns and trends in the data, identifying anomalies and unusual readings, making predictions based on the available data, predicting and preventing adverse events, and developing personalized treatment plans for patients. For instance, Iwendi et al. [6] developed an IoMT-based diet recommendation system, Sayeed et al. [7] developed an IoMT-based seizure detection model, Khan et al. [8] developed an IoMT model for elderly monitoring, Siddiqui et al. [9] developed an IoMT-based cancer detection model, Datta Gupta et al. [10] developed a histopathological classification model using machine learning techniques.

Despite the potential of machine learning techniques in IoMT to revolutionize healthcare, several challenges remain. The conventional machine learning models in the IoMT domain are static in that they were trained on some datasets and are being used for real-time inferencing data. This approach does not consider the patient's recent health-related data, i.e., the data that the IoMT devices have provided, since the models are not trained on the recent data. The models need to be aware of the latest health updates of the patient. Further, in the conventional machine learning models paradigm, the models must be re-trained again, even to incorporate a few sets of additional samples. Since the training of the conventional machine learning models generally happens on cloud platforms, there are also risks to security and privacy.

In this work, we propose an IoMT-based edge-enabled incremental learning framework for intelligent diagnosis in smart healthcare. Incremental learning [11] is a machine learning paradigm that allows a model to learn continuously from new data. Incremental learning algorithms are adaptive, flexible, and efficient. The model can learn from new data, adapt to new patterns, and improve its accuracy over time. It is ideal for use in a smart healthcare environment where patient data is constantly generated. The proposed incremental learning framework can handle the data in real-time in an efficient way. The proposed model can also learn data samples one by one. The proposed incremental framework requires less computational resources and storage space than traditional models because it processes and stores only relevant and new data. Thus, the proposed IoMT framework is more cost-effective than the conventional machine learning-based frameworks. On the other hand, conventional models require significant computational resources and storage space to handle large datasets and complex models. Another significant contribution of incorporating incremental learning is privacy: the proposed incremental models are trained locally on edge devices, which can help preserve the confidentiality of sensitive data.

We propose a three-layer edge-based incremental learning framework. The first and last layers are used for data collection and application deployment, respectively, in line with the existing works [7], [8]. The novel second layer (edge layer) allows incremental training and inferencing at the edge. To this extent, we proposed two algorithms for incremental classification and anomaly detection on the preprocessed IoMT data, inspired by Aggregated Mondrial Forests (AMF) and Half-Space Trees (HST) algorithms, respectively. Addressing the issue of excessive features, which might increase the computational burden on incremental models, we proposed a novel feature selection algorithm based on the concepts of bijective soft sets, Shannon entropy, and TOPSIS. Detailed experimentation shows that the proposed algorithms perform better than the other incremental algorithms by a good margin. More importantly, the proposed algorithms achieve significant improvement (>13%) in terms of accuracy, and they also reduce the computational time by more than 80x, compared to the conventional batch-based machine learning algorithms.

A. CONTRIBUTIONS

The main contributions of the paper can be highlighted below:

i) We propose a novel edge-enabled IoMT-based incremental learning framework that addresses several issues of the conventional machine learning models.
ii) We propose an advanced feature selection algorithm based on bijective soft sets, Shannon entropy, and TOPSIS.
iii) We propose two incremental algorithms for classification and anomaly detection inspired by Aggregated Mondrial Forests (AMF) and Half-Space Trees (HST) algorithms.
iv) We extensively evaluate the proposed algorithms, comparing their performance and computational time concerning the batch-learning-based algorithms.

B. ORGANIZATION

The rest of the paper is arranged as follows. Section II reviews the works related to the proposed framework. Section III describes the overall workflow of the proposed IoMT framework, the proposed feature selection algorithm, and the proposed incremental algorithms for classification and anomaly detection. Section IV discusses the performance of the proposed algorithms in terms of different metrics. Finally, Section V concludes the work.

II. RELATED WORK

This section describes the works related to intelligent diagnosis in smart healthcare systems and identifies the research gaps.

A. IoMT FOR INTELLIGENT DIAGNOSIS IN HEALTHCARE

The article [12] proposes the MSSO-ANFIS framework as a culmination of the Adaptive Neuro-Fuzzy Inference System (ANFIS) and Modified Salp Swarm Optimization (MSSO) algorithm for investigating and identifying the key characteristics of heart disease using IoMT devices. The framework collects patient data, which is stored and transmitted to the cloud constantly and in real time. The data is passed to the hospital administration to identify the disease after a prediction from the ML framework. Levy crow search algorithm [13] is used for feature selection in the proposed framework. The authors report an accuracy of 99.45% and a precision of 96.54%.

B. INCREMENTAL LEARNING-BASED ANOMALY DETECTION

In the paper [14], the authors addressed the problem of average monitoring performance and feature extraction due to the problem in conventional batch processes of ignoring mutual correlation among process variables and autocorrelation. The issue has been addressed by proposing a Dynamic Hidden Variable Fuzzy Broad Neural network(DHVFBN). Going into detail, Slow Feature Analysis (SFA) addresses non-linearity and captures dynamic features. The model's Quick reconstruction and updating are done using Fuzzy Broad Learning System (FBLS), which reduces the computation time as the model need not retrain after new faults are added. In the paper [15], the authors proposed Online Incremental Learning Algorithm (OILA), which uses a regression-based approach to predict health parameters and incorporates a feedback mechanism to increase the accuracy. Alerts are generated whenever an anomaly is seen in the patient's health parameters. It aims at solving a current challenge with the anomaly detection algorithms-not being able to process the data incrementally and hence needing to improve at anomaly detection and predicting anomaly at a correct instance. The proposed algorithm is compared with the Kalman Filter. Lastly, they have validated their algorithm with real-time health parameter data sets.

C. INCREMENTAL LEARNING-BASED CLASSIFICATION

In [16], a pattern classification system is presented in which classifier learning and feature extraction are simultaneously done online and in one pass. It was achieved by combining some classifier models and incremental principal component analysis (IPCA). The problem the authors recognised with the approach mentioned above and worked on is because of the limitations of IPCA; the training samples had to be learned individually. They propose to address this problem by chunk IPCA, which takes many samples to process at a time. In order to discuss the scalability of IPCA in one-pass incremental learning situations, the classification performance with large-scale datasets was evaluated. The results show that chunk IPCA reduces training time in comparison to IPCA. The authors also studied the influence of the size of given chunk data and the size of initial training data on learning time and classification accuracy. In [17], the authors focus on the forgetfulness aspect in incremental learning, leading to a significant performance drop. The paper surveys class-incremental learning for image classification and performs substantial experiments on thirteen class implementation methods. Another set of experiments included a comparison of class-implemental on multiple large-scale image classifications.

III. PROPOSED METHODOLOGY

This section describes the proposed IoMT-based incremental learning framework in detail. Firstly, we will explain the overall workflow of the framework, followed by a clear description of its components.

A. OVERALL WORKFLOW

Figure 1 represents the proposed framework. The proposed framework broadly consists of three layers - the device, edge, and application layers. The device layer consists of various IoMT devices like SPO2 sensors, smart watches measuring blood pressure and heart beat rate, EEG devices, etc., which monitor the patient and send the signals to the edge layer.

The edge layer is the most essential in the proposed framework. In the traditional frameworks for smart diagnosis in healthcare, the machine learning models are deployed on the cloud [18]. However, such deployments have latency and security issues. Addressing this issue, several researchers tried to train machine learning models at the edge devices [19], [20]. Although edge-based deployment addresses the concerns with cloud-based deployment, it has a notable drawback: the machine learning models deployed at the edge must have a lower memory footprint and computational requirements, which is often not the case, given the model size and performance tradeoff. We propose deploying incremental machine learning models on edge devices in this work. Given the lower memory footprint and lesser computational requirements of incremental machine learning models, the issues in conventional edge-based deployment will not be raised here. Besides the lower memory footprint, edge-based deployment has additional benefits in the proposed framework.

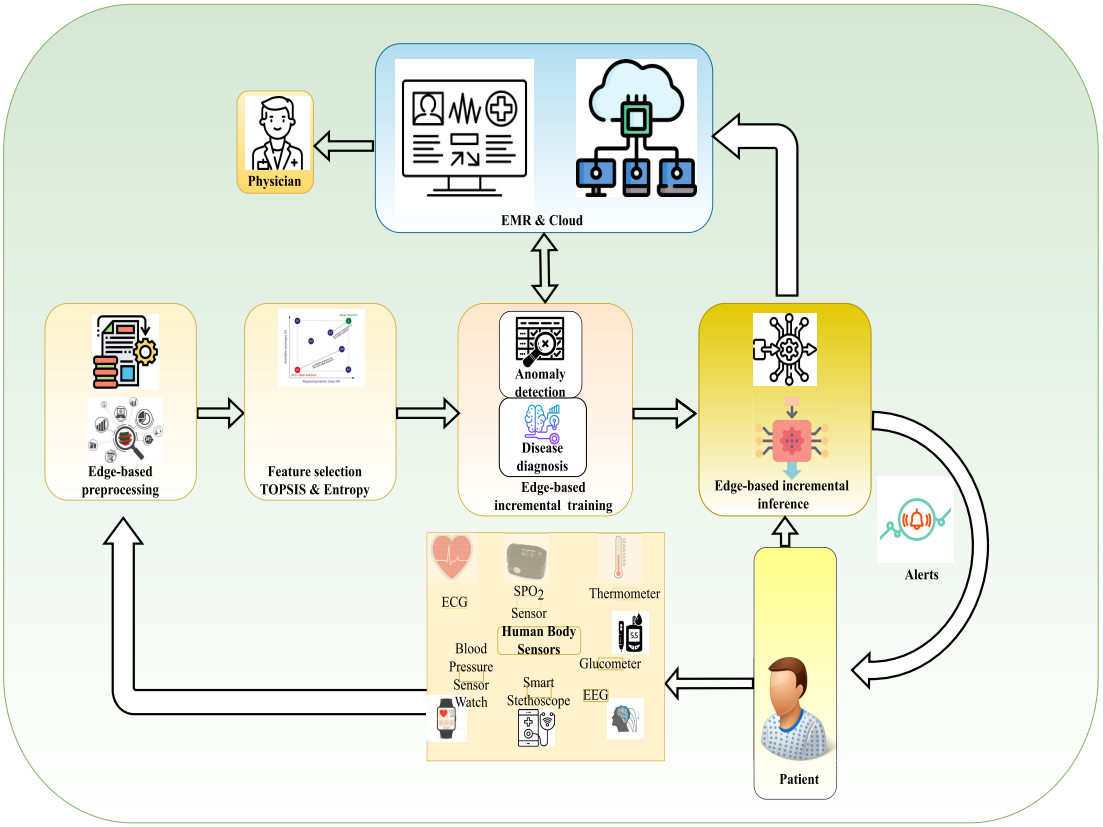


FIGURE 1. **Proposed methodology.**

i) The proposed framework has an anomaly detection machine learning model to identify the anomalies in the IoMT data of patients. Given the critical nature of anomaly detection tasks, detecting and responding to potential anomalies in less time is important. The edge deployment reduces the latency and is hence helpful.
ii) Given the concern of data security in healthcare, with the help of edge-based deployment, the sensitive data can be stored on the local network, thus reducing the risks of cyber-attacks and data breaches.
iii) The proposed system also includes an incremental learning-based inference system. The edge deployment also reduces the need to transmit the data to the cloud while improving the speed and accuracy of the model.

The edge layer consists of four broad components: edge-based data preprocessing and analytics, feature selection mechanism, edge-based incremental learning model, and edge-based incremental inference. The data from the patient's IoMT sensors will be preprocessed and analyzed for insights into the edge preprocessing component. Then, the data will be passed to the edge incremental learning component after feature selection. The edge incremental learning component consists of two online machine-learning models for anomaly detection and disease diagnosis. The proposed algorithms for these models are presented in detail later. After the models are trained, the weights and other parameters are stored on the cloud for retraining or inference. In the proposed system, the anomalies in the data and the disease diagnosis are inferred in an interval fashion. The anomaly detection inference happens more frequently than disease diagnosis for apparent reasons. The IoMT data will be preprocessed and passed to the inference component, which infers and alert the patient whenever required. The inferences will also be stored on the cloud for use by physicians.

The final application layer consists of a secure cloud storage platform. The cloud also consists of the patient's electronic medical records, which organize the information regarding the patient systematically. Apart from the EMRs, the parameters and hyper-parameters of the edge-trained machine learning models will also be stored on the cloud. With the access permissions from the patient, a physician can access the EMR information of the patient, which aids in the patient's treatment.

B. FEATURE SELECTION EMPLOYING BIJECTIVE SOFT SETS, SHANNON ENTROPY AND TOPSIS

In the conventional IoMT setting, several features can be surged due to an increase in the available IoMT sensor readings from the patient. All these features may need to be more helpful in performing classification and anomaly detection, thus raising the need for a feature selection algorithm. Feature selection is particularly important in the proposed edge-based

incremental setting, where one of the main goals is to reduce the computational burden. In this section, we describe the components of the proposed feature selection algorithm, followed by the implementation details.

1) SOFT SETS

A soft set is a mathematical concept used to represent imprecise or incomplete information in a set. This extends the concept of a traditional set by introducing a degree of uncertainty in the membership of the elements.

Definition 1: If U represents the universe set, E represents the parameter set chosen for analysis and $P(U)$ represents the power set of U, then pair (F, E) is said to be a soft set over U, where F is a function that maps elements from E to subsets of U. This can be shown as:

$$F : E \rightarrow P(U)$$

.

Definition 2: A soft set (F, E) over the universal set U is considered to be a bijective soft set if it meets the following criteria [21]:

1)

$$\bigcup_{e \in E} F(e) = U$$

2) For any two parameter values $e_i, e_j \in E, e_i \neq e_j$

$$F(e_i) \cap F(e_j) = \emptyset$$

2) SHANNON ENTROPY

The Shannon entropy tells about the level of uncertainty in information. This uncertainty can be expressed using probability theory. The function H, proposed by Shannon, fulfills the specified conditions for the values of p_i in the estimated joint probability distribution P.

$$H(P) = -p_i \log(p_i)$$

The conditions are as follows:

1) H is a continuous function that is always positive.
2) When all p_i are equal and defined as $\frac{1}{n}$, H exhibits a monotonic increase with respect to n.
3) When considering any value of n greater than or equal to 2, the function $H(p_1, \ldots, p_n)$ can be expressed as the sum of two terms. The first term is $H(p_1 + p_2, p_3, \ldots, p_n)$, and the second term is $(p_1 + p_2)H\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right)$.

Shannon showed that H is the only function that satisfies all these conditions.

Shannon Entropy is used to capture the preferences of network security experts. This concept is helpful for calculating weights of parameter values [22], which tells about their relative importance. The projection value is calculated, which is used to calculate entropy, then divergence, and finally, weight for the parameter values.

3) TOPSIS (TECHNIQUE FOR ORDER OF PREFERENCE BY SIMILARITY TO IDEAL SOLUTION)

It is a multi-criteria decision analysis technique that compares various alternatives and ranks them based on their distance from ideal and non-ideal solutions. The optimal choice is the alternative that minimizes the distance from the ideal solution while maximizing the distance from the non-ideal solution. TOPSIS procedure includes several key steps. Firstly, the alternatives and criteria are identified and put into a matrix. Then, we normalize the matrix and calculate the weighted normalized decision matrix. After this, we determine ideal and non-ideal solutions and calculate the relative closeness of the alternatives to the ideal solution. The alternative with maximum relative closeness is taken as the optimal choice.

4) IMPLEMENTATION

This section applies the proposed methodology for the breast cancer dataset.

1) A set of effective features is selected by us as follows: $EF = [EF_1, EF_2, EF_3, EF_4, EF_5]$,where

   EF_1 = marginal adhesion
   EF_2 = epithelial cell diameter
   EF_3 = blandness of nuclear chromatin
   EF_4 = infrequent mitoses
   EF_5 = uniformity of epithelial cell size

   These features selected can also be termed parameters or attributes. The following values are assigned to the features selected. These values represent target values for network security requirements.

   $EF_1 = \{x_{11}, x_{12}, x_{13}\}$ = {Low, Medium, High}
   $EF_2 = \{x_{21}, x_{22}, x_{23}\}$ = {Poor, Good, Very Good}
   $EF_3 = \{x_{31}, x_{32}, x_{33}\}$ = {Poor, Good, Very Good}
   $EF_4 = \{x_{41}, x_{42}, x_{43}\}$ = {Low, Medium, High}
   $EF_5 = \{x_{51}, x_{52}, x_{53}\}$ = {Low, Medium, High}

2) Five combinations are formed by combining parameter values from each useful feature. In this context, a suitable combination refers to selecting parameter values from the useful features representing a specific case (an alternative). These combinations are termed functional concepts. The union of these concepts form the universe set U.

   $$U = \{EF_1, EF_2, EF_3, EF_4, EF_5\}$$

   The functional concepts are given as follows:

   $FC_1 = \{x_{12}, x_{21}, x_{33}, x_{42}, x_{53}\}$
   $FC_2 = \{x_{11}, x_{23}, x_{31}, x_{42}, x_{52}\}$
   $FC_3 = \{x_{13}, x_{22}, x_{32}, x_{43}, x_{51}\}$
   $FC_4 = \{x_{11}, x_{22}, x_{31}, x_{41}, x_{51}\}$
   $FC_5 = \{x_{12}, x_{23}, x_{31}, x_{42}, x_{52}\}$

3) The valuable features can be expressed as soft sets, which assist the network security expert in determining which functional concept fulfills a specific target value.

   $(F_1, EF_1) = \{F_1(x_{11}), F_1(x_{12}), F_1(x_{13})\}$
   $(F_2, EF_2) = \{F_2(x_{21}), F_2(x_{22}), F_2(x_{23})\}$

TABLE 1. Preference decision matrix.

	x_{11}	x_{12}	x_{13}	x_{21}	x_{22}	x_{23}	x_{31}	x_{32}	x_{33}	x_{41}	x_{42}	x_{43}	x_{51}	x_{52}	x_{53}
NS_1	0.5	0.5	0.5	0.5	0.2	0.5	0.7	0.7	0.9	0.7	0.5	0.7	0.9	0.5	0.2
NS_2	0.9	0.2	0.7	0.9	0.2	0.5	0.7	0.2	0.5	0.2	0.2	0.7	0.2	0.5	0.9
NS_3	0.7	0.2	0.9	0.7	0.5	0.9	0.5	0.2	0.9	0.7	0.5	0.9	0.5	0.2	0.7

TABLE 2. Projection(p_{ij}), Entropy(*Ent*), Divergence(*Div*) and Weight(W).

	x_{11}	x_{12}	x_{13}	x_{21}	x_{22}	x_{23}	x_{31}	x_{32}	x_{33}	x_{41}	x_{42}	x_{43}	x_{51}	x_{52}	x_{53}
NS_1	0.238	0.556	0.238	0.238	0.222	0.263	0.368	0.636	0.391	0.438	0.417	0.304	0.562	0.417	0.111
NS_2	0.429	0.222	0.333	0.429	0.222	0.263	0.368	0.182	0.217	0.125	0.167	0.304	0.125	0.417	0.5
NS_3	0.333	0.222	0.429	0.333	0.556	0.474	0.263	0.182	0.391	0.438	0.417	0.391	0.312	0.167	0.389
Ent	0.975	0.906	0.975	0.975	0.906	0.962	0.989	0.826	0.970	0.895	0.936	0.993	0.862	0.936	0.872
Div	0.025	0.094	0.025	0.025	0.094	0.038	0.011	0.174	0.030	0.105	0.064	0.007	0.138	0.064	0.128
W	0.025	0.047	0.013	0.013	0.050	0.021	0.006	0.100	0.019	0.069	0.045	0.005	0.102	0.053	0.111

TABLE 3. NS_1.

	x_{12}	x_{22}	x_{31}	x_{43}	x_{52}	WUV
FC_1	1	0	0	0	0	0.0467
FC_2	0	0	1	0	1	0.0060
FC_3	0	1	0	1	0	0.0182
FC_4	0	1	1	0	0	0.0193
FC_5	1	0	1	0	1	0.0527
W	0.0467	0.0132	0.0060	0.0049	0.0528	

TABLE 4. NS_2.

	x_{11}	x_{23}	x_{32}	x_{41}	x_{53}	WUV
FC_1	0	0	0	0	1	0.1113
FC_2	1	1	0	0	0	0.0461
FC_3	0	0	1	0	0	0.1004
FC_4	1	0	0	1	0	0.0933
FC_5	0	1	0	0	0	0.0215
W	0.0246	0.0215	0.1004	0.0687	0.1113	

TABLE 5. NS_3.

	x_{13}	x_{21}	x_{33}	x_{42}	x_{51}	WUV
FC_1	0	1	1	1	0	0.0773
FC_2	0	0	0	1	0	0.0451
FC_3	1	0	0	0	0	0.0130
FC_4	0	0	0	0	1	0.1020
FC_5	0	0	0	1	0	0.0451
W	0.0130	0.0132	0.0190	0.0451	0.1020	

TABLE 6. NIS and IS for each NS.

	x_i^*	x_i^-
NS_1	0.0527	0.0060
NS_2	0.1113	0.0215
NS_3	0.1020	0.0130

$$(F_3, EF_3) = \{F_3(x_{31}), F_3(x_{32}), F_3(x_{33})\}$$
$$(F_4, EF_4) = \{F_4(x_{41}), F_4(x_{42}), F_4(x_{43})\}$$
$$(F_5, EF_5) = \{F_5(x_{51}), F_5(x_{52}), F_5(x_{53})\}$$

F is a function that maps elements from EF to subsets of U. The mapping is given as follows: $F_1(x_{11}) = FC_2, FC_4$; $F_1(x_{12}) = FC_1, FC_5$; $F_1(x_{13}) = FC_3$; $F_2(x_{21}) = FC_1$; $F_2(x_{22}) = FC_3, FC_4$; $F_2(x_{23}) = FC_2, FC_5$; $F_3(x_{31}) = FC_2, FC_4, FC_5$; $F_3(x_{32}) = FC_3$; $F_3(x_{33}) = FC_1$; $F_4(x_{41}) = FC_4$; $F_4(x_{42}) = FC_2, FC_5$; $F_4(x_{43}) = FC_3$; $F_5(x_{51}) = FC_3, FC_4$; $F_5(x_{52}) = FC_2, FC_5$; $F_5(x_{53}) = FC_1$.

All the above soft sets fulfill the conditions required for being a bijective soft set. As an example, in the softset (F_1, EF_1), the combination of all the soft sets of (F_1, EF_1) forms the common universe U or $\bigcup_{x_{1j} \in EF_1} F_1(x_{1j}) = U$. For the second condition, taking any two parameter values, suppose $x_{11}, x_{12} \in EF_1$ where $x_{11} \neq x_{12}$, it holds that $F1(x_{11}) \cap F1(x_{12}) = \emptyset$ indicating that there is no intersection between the soft sets associated with different parameter values.

4) The network security (NS) experts were required to give preferences to different features' values. These preferences will be used in Shannon Entropy. The available options for preferences include Low (0.2), Medium (0.5), High (0.7), and Very high (0.9). The preference decision matrix is shown in Table 1, which collects all the preferences given by NS experts.
5) We then calculate the projection value (p_{ij}) for each parameter value x_{ij} as:

$$p_{ij} = \frac{f_{ij}}{\sum_{i=1}^{m} f_{ij}}$$

The entropy values e_j for each parameter value x_{ij} is calculated as:

$$e_j = -k \sum_{i=1}^{m} p_{ij} \ln(p_{ij}), \text{where } k = (\ln(m))^{-1}$$

The degree of divergence d_j is calculated as:

$$d_j = 1 - e_j$$

For each parameter value x_{ij} the weight W_j is calculated as:

$$W(x_{ij}) = \sum_{k=1}^{n} \frac{d_j}{d_k}$$

All these values are tabulated in Table 2.

TABLE 7. Separation measure of each NS from NIS and IS.

	$D^*_{1,k}$	$D^-_{1,k}$	$D^*_{2,k}$	$D^-_{2,k}$	$D^*_{3,k}$	$D^-_{3,k}$
FC_1	0.000036	0.001654	0.000000	0.008063	0.000611	0.004128
FC_2	0.002181	0.000000	0.004254	0.000604	0.003246	0.001024
FC_3	0.001195	0.000147	0.000118	0.006234	0.007916	0.000000
FC_4	0.001121	0.000175	0.000324	0.005154	0.000000	0.007916
FC_5	0.000000	0.002181	0.008063	0.000000	0.003246	0.001024

TABLE 8. Combined separation.

	D^*_k	D^-_k
FC_1	0.025446	0.117667
FC_2	0.098388	0.040349
FC_3	0.096068	0.079880
FC_4	0.038010	0.115088
FC_5	0.106343	0.056611

TABLE 9. Relative closeness of FC.

	C^*_k
FC_1	0.822198
FC_2	0.290829
FC_3	0.454000
FC_4	0.751726
FC_5	0.347407

6) The network security experts select different parameter values which form the requirement sets. The best functional concept will be chosen based on these requirement sets. The sets are as follows:

$$NS_1 = \{x_{12}, x_{22}, x_{31}, x_{43}, x_{52}\}$$
$$NS_2 = \{x_{11}, x_{23}, x_{32}, x_{41}, x_{53}\}$$
$$NS_3 = \{x_{13}, x_{21}, x_{33}, x_{42}, x_{51}\}$$

Tables 3, 4 and 5 represent these sets as soft sets. The Weighted Utility Value (WUV) is also computed in these tables which tells about the effectiveness of each functional concept for a given NS_i. $WUV_{i,k}$, weighted utility value for a requirement set of NS_i and functional concept k is computed as:

$$WUV_{i,k} = \sum_j d_{kj}, \text{ where } d_{kj} = W(x_{mn}) \times h_{kj}$$

x_{mn} represents a column iterated by j, W is the weight of each column and h_{kj} is used for accessing values in the softset table.

7) The ideal solution (x^*_i) and non-ideal solution (x^-_i) is calculated for each requirement set NS_i as:

$$x^*_i = \max(WUV_{i,k})$$
$$x^-_i = \min(WUV_{i,k})$$

These are shown in Table 6.

8) For each NS_i, the separation measure from an ideal solution and a non-ideal solution is calculated as:

$$D^*_{i,k} = (WUV_{i,k} - x^*_i)^2$$
$$D^-_{i,k} = (WUV_{i,k} - x^-_i)^2$$

These are shown in Table 7.
The combined separation measure is shown in Table 8. It is calculated as:

$$D^*_k = \sqrt{\sum_{i=1}^{i=m} D^*_{i,k}}$$
$$D^-_k = \sqrt{\sum_{i=1}^{i=m} D^-_{i,k}}$$

9) The relative closeness between the concepts FC_k and the ideal solution is calculated as:

$$C^*_k = \frac{D^-_k}{D^-_k + D^*_k}$$

These values are shown in Table 9. Now, the cases are ranked according to their C^*_k values in descending order. Ranking is given as

$$\{FC_1, FC_4, FC_3, FC_5, FC_2\}$$

The best concept is the one with the maximum value of C^*_k. This particular alternative will exhibit the least distance from the ideal solution and the greatest distance from the non-ideal solution. Based on the above requirements, the functional concept FC_1 emerges as the optimal choice. The relative closeness for FC_1 is 0.822 which is highest compared to others. FC_1 is composed of x_{12}, x_{21}, x_{33}, x_{42} and x_{53} which can be inferred as medium, poor, very good, medium and high respectively.

C. PROPOSED INCREMENTAL ALGORITHM FOR CLASSIFICATION

We proposed an incremental learning classification algorithm inspired by the AMF algorithm for real-time IoMT data classification [23]. The proposed algorithm is an ensemble learning method that combines the Mondrian Forest algorithm [24] with an aggregation scheme. This combination improves the performance of the algorithm for online learning tasks.

Algorithm 1 Proposed Incremental Learning Algorithm for Classification

Input : Data stream D, number of trees $n_{estimators}$, step-size for the aggregation weights *step*, use of aggregation in trees *use_aggregation*, regularization level of class frequencies *dirichlet*, use of pure node splitting *split_pure*, subsampling fraction f and random seed *seed*.

Output: Ensemble of $n_{estimators}$, Mondrian trees $T = \{T_1, \ldots, T_{n_{estimators}}\}$ with aggregation weights $w_{1:T}$.

Initialize $w_{1:T}$ to 1

for $t = 1$ **to** $n_{estimators}$ **do**

  $D_t \leftarrow$ random subset of D with fraction f; $T_t \leftarrow MondrianTree(D_t, w_t, dirichlet, split\_pure)$;

  if *use_aggregation* **then**

    $$w_t \leftarrow \frac{w_{t-1} \exp(-step * \hat{L}(T_t, D_t))}{\sum_{i=1}^{t} w_i \exp(-step * \hat{L}(T_i, D_i))}$$

  end

end

Function *MondrianTree(D, w, dirichlet, split_pure)*:

  if $|D| = 1$ **then**

    return *Leaf node with a class of the single instance in D*

  end

  else if *all instances in D have the same class* **then**

    return *Leaf node with the common class*

  end

  else

    Select feature j and split value v

    $D_L \leftarrow \{x \in D : x_j \leq v\}$; $D_R \leftarrow \{x \in D : x_j > v\}$

    $T_L \leftarrow MondrianTree(D_L, w, dirichlet, split\_pure)$

    $T_R \leftarrow MondrianTree(D_R, w, dirichlet, split\_pure)$

    $p_L \leftarrow \frac{|D_L|}{|D|}$; $p_R \leftarrow \frac{|D_R|}{|D|}$

    return *Internal node with split feature j, split value v, child nodes T_L and T_R, and class probabilities p_L and p_R.*

  end

Algorithm 1 and Figure 2 presents the proposed algorithm. The algorithm creates a set of Mondrian trees on a data stream. Each tree in the ensemble is constructed on a randomly selected subset of the data with a fraction f and has a weight w_t. The weights of the trees are updated iteratively based on the performance of each tree. The goal is to learn a set of Mondrian trees that minimize the data stream's loss function value. The algorithm starts by initializing the weights of all trees to 1. Then, for each tree T_t from 1 to $n_{estimators}$, the algorithm randomly selects a subset of the data with a fraction f to build the tree. The tree is constructed using the MondrianTree function that recursively partitions the data into smaller subsets based on a selected feature and a split value.

If aggregation is enabled, the weight w_t of the current tree is updated according to the performance of all previous trees in the ensemble. The update rule is a weighted exponential

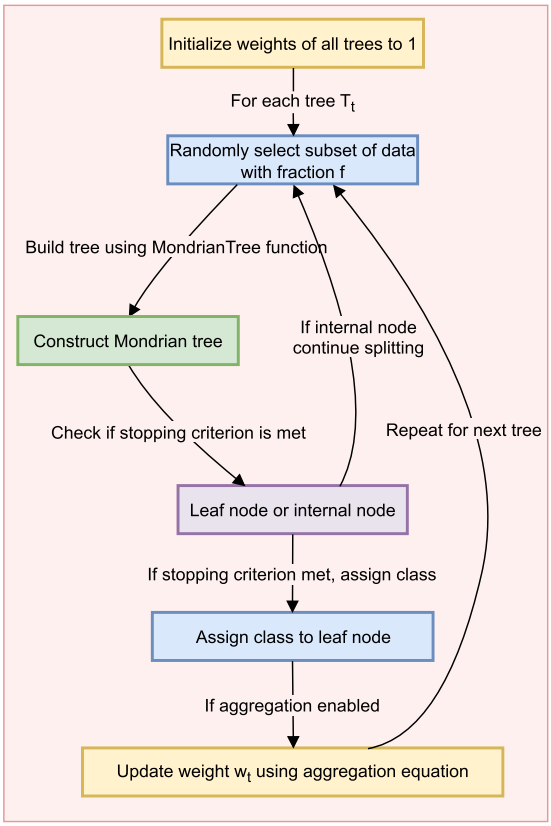


FIGURE 2. Flow chart of the proposed algorithm for classification.

function penalizes trees with high loss values. The weight of the current tree is updated using the following equation:

$$w_t = \frac{w_{t-1} \exp(-step * \hat{L}(T_t, D_t))}{\sum_{i=1}^{t} w_i \exp(-step * \hat{L}(T_i, D_i))}$$

where $\hat{L}(T_i, D_i)$ is the estimated loss of tree T_i on data D_i, and *step* is the step-size for the aggregation weights.

The *MondrianTree*function (in the second part of Algorithm 1) builds a decision tree, which partitions the data into a hierarchical structure that preserves the temporal order of the stream. The tree is constructed recursively by selecting a feature and then a split value that maximizes the reduction in the expected loss. The expected loss is a weighted sum of the losses of the child nodes. The weights are the fraction of instances that belong to each child node. The process is continued until a stopping criterion is met, when all instances in a node belong to the same class or the node contains only one instance. The resulting tree is either a leaf node with the class of the single instance in the subset or an internal node with a split feature, split value, child nodes, and class probabilities.

D. PROPOSED INCREMENTAL LEARNING FOR ANOMALY DETECTION

We proposed an incremental learning anomaly detection algorithm inspired by the Streaming Half-Space Trees

TABLE 10. Comparison of proposed algorithm with batch algorithms for classification.

Model	Accuracy	Time(s)
SVM	75.97	10.749
Decision Tree	74.02	3.257
Naive Bayes	74.02	2.845
Proposed model	**87.63**	**1.299**

(HS-Trees) algorithm [25] for real-time anomaly detection in the IoMT data. The streaming data coming in as input is generally of high velocity and volume; therefore, traditional batch-based learning methods can become computationally infeasible due to their high computational and memory requirements. Our proposed model particularly addresses streaming data.

Algorithm 2 presents the proposed incremental algorithm for anomaly detection. It operates on a data stream D with a given window size γ and a specified number of HS-Trees n. The aim is to compute an anomaly score scr for a streaming instance x. The algorithm begins by creating a specified number of HS-Trees. The *MassUpdate* function is invoked for each instance x in the first γ instances of the stream that updates the mass of the corresponding nodes in the trees. If it's the reference window, the algorithm increments the reference mass ($Node.r$), otherwise, it increments the latest mass ($Node.l$). If the current node's depth ($Node.k$) is less than the maximum depth, the function recursively calls itself for the next level node that x traverses. This update in mass for the first γ instances creates an initial reference mass profile for the HS-Trees. The reference mass profile is used to infer anomaly scores of new data coming in the latest window.

The *Score* function computes the anomaly score for an instance x in an HS-Tree T. It traverses x from the root of T to a terminal node $Node_t$ and returns $Node_t.r \times 2^{Node_t.k}$, where $Node_t.k$ is the depth level of the terminal node and $Node_t.r$ is the mass of the terminal node.

In the algorithm's main loop, a streaming point x is received and scr is set to 0. For each tree T in the HS-Trees, the *Score* function is called to compute the anomaly score for x in T, and the *MassUpdate* function is invoked with the reference window set to *false*. This updates the latest mass in the corresponding nodes. After accumulating the tree scores, the algorithm reports scr as the anomaly score for x.

After scoring a batch of data points, the algorithm updates the model. It transfers the latest mass values from the nodes in the latest window to the nodes in the reference window. This ensures that the reference window always contains the most up-to-date mass profile for scoring the next batch of data points. The count variable c keeps track of the number of instances processed. When c reaches γ, the model is updated by transferring the non-zero latest mass l to each node's reference mass r. Then, the latest mass l is reset to 0 for nodes with non-zero mass l and the c is reset to 0.

Algorithm 2 Proposed Incremental Learning Algorithm for Anomaly Detection

Input: Data Stream D, Window Size γ, Number of HS-Trees n

Output: scr – anomaly score for each streaming instance x

```
Construct n HS-Trees and for each tree T do
    for each instance x in the first γ instances of the stream do
        MassUpdate(x, T.root, true)
    end
end
Save this initial reference mass profile for each HS-tree T.

Function MassUpdate(x, Node, referenceWindow):
    if referenceWindow then
        Increment Node.r by 1
    end
    else
        Increment Node.l by 1
    end
    if Node.k < MaximumDepth then
        Let Node* represent the subsequent level of Node that x visits
        MassUpdate(x, Node*, referenceWindow)
    end

Function Score(x, T):
    Traverse x from the root of T to a terminal node (Node_t)
    return Node_t.r × 2^(Node_t.k)
     where Node_t.k represents the depth level of the terminal node having mass Node_t.r

Initialize count c to 0
while data stream flows do
    Receive the next streaming instance x, set scr to 0
    for each tree T in HS-Trees do
        scr ← scr+ Score(x, T)
        MassUpdate(x, T.root, false)
    end
    Return scr {the anomaly score for instance x}
    Increment c by 1
    if c = γ then
        for each node with non-zero mass r or l do
            Set Node.r to Node.l;
        end
        for each node with non-zero mass l do
            Reset Node.l to 0;
        end
        Reset c to 0
    end
end
```

IV. EXPERIMENTATION AND PERFORMANCE EVALUATION

In this section, the conducted experiments along with the comparisons with the baseline and existing state-of-the-art

TABLE 11. Performance comparison of the incremental models for classification.

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression(baseline)	73.96	61.49	**67.91**	64.54
Passive Aggressive classifier [26]	58.07	39.37	37.31	38.31
ALMA [27]	66.67	51.79	64.93	57.62
ARF [28]	73.01	65.00	48.69	55.67
Proposed model	**87.63**	**92.31**	67.16	**77.75**

TABLE 12. Comparison of proposed algorithm with batch algorithms for anomaly detection.

Model	Accuracy	Time(s)
Elliptic Envelope	74.52	0.458
OneClassSVM	84.04	0.016
Isolation Forest [29]	95.46	0.082
Proposed Model	**97.22**	**0.001**

models are described, beginning with a brief description of the datasets and evaluation metrics used.

A. DATASETS AND EXPERIMENTAL SETUP

Diabetes data set [30] is used for classification. This dataset encompasses critical attributes, including pregnancies, glucose levels, and diabetes pedigree functions. It has 768 instances, eight features and two classes (whether the patient has diabetes or not). All features are numerically valued. We use the Breast Cancer dataset [31] for anomaly detection. It consists of 683 instances, nine features and two classes which are benign and malignant. The features are cytological characteristics like clump thickness, marginal adhesion, infrequent mitoses, uniformity of epithelial cell size, epithelial cell diameter, bare nuclei, normal nucleoli, uniformity of cell shape and blandness of nuclear chromatin.

In our research, we implemented the machine learning models using Python programming language. Many open-source libraries were used, including Numpy, Pandas, Scikit-learn and Matplotlib. Example taken on a system with: Intel Core i5-9300H 2.40GHz processor, 8GB of RAM and NVIDIA GeForce GTX 1650 graphics card. We have used accuracy, ROCAUC, Logloss, and weighted F1-score as metrics for evaluating the models. Apart from these metrics, we have also measured the time taken by the models using the time module in Python.

B. INCREMENTAL CLASSIFICATION

We compare the performance of the model proposed in the table 11 with other basic incremental algorithms and state-of-the-art incremental algorithms [26], [27], [28]. The proposed model obtains the best accuracy, precision, and F1-score values. The recall score of the proposed models is also quite close to the best recall score. We evaluate the performance and duration of the proposed model compared to the aggregated algorithms in the Table 10.

As can be seen from the table, the proposed model has the best accuracy on the Diabetes dataset and takes the least training time compared to other batch learning algorithms. SVM takes a lot of time to train. The accuracy results of the Decision Tree and Naive Bayes algorithm are similar. Overall, our model gives the best results in terms of accuracy and time taken in comparison to other batch models.

Figure 3 presents the plots of ROCAUC, log loss, accuracy and weighted F1-score of the proposed incremental algorithm for classification. From Figure 3a, it can be seen that the graph starts to saturate at around 100 data points, after which it starts to increment slowly, suggesting that as we increase the size of the dataset, the ROCAUC value will increase, which establishes the model's better performance. From figure 3b, it can be observed that the graph starts to saturate after around 100 data points. Then it slowly decrees, suggesting that as the dataset increases, the log loss decreases as the model becomes better at predicting. From Figure 3d, it can be inferred that the model's weighted F1-score stabilizes around 200 data points and very acutely increases afterward. Figure 3c shows that when the model is in its initial stage, accuracy is less because the model is not trained with much data, but as the size of the Dataset increases, the model learns how to predict more precisely, and accuracy starts to increase.

Figure 4 compares the proposed incremental algorithm with the batch learning algorithms for classification in terms of accuracy and time. As can be seen from the figure 4a, the proposed model performs best in terms of accuracy, followed by the ensemble learning model of Naive Bayes, Support Vector Machine and Decision Tree. The proposed model has low accuracy initially, but as the model learns with each iteration, accuracy improves and outperforms other models. From the Figure 4b it can be judged that our proposed model is is more computationally efficient compared to other batch processing methods.

C. INCREMENTAL ANOMALY DETECTION

Table 13 presents the performance comparison of the proposed model, a baseline incremental algorithm and state-of-the-art incremental algorithms [27], [32]. Our proposed model outperforms every other model in all the metrics with an accuracy of 97.22%. The proposed model is better in terms of accuracy than One class SVM with a Quantile filter [32] by a difference of 1.68%. The One-Class SVM with Quantile filter is the second-best model with an accuracy of 95.90%.

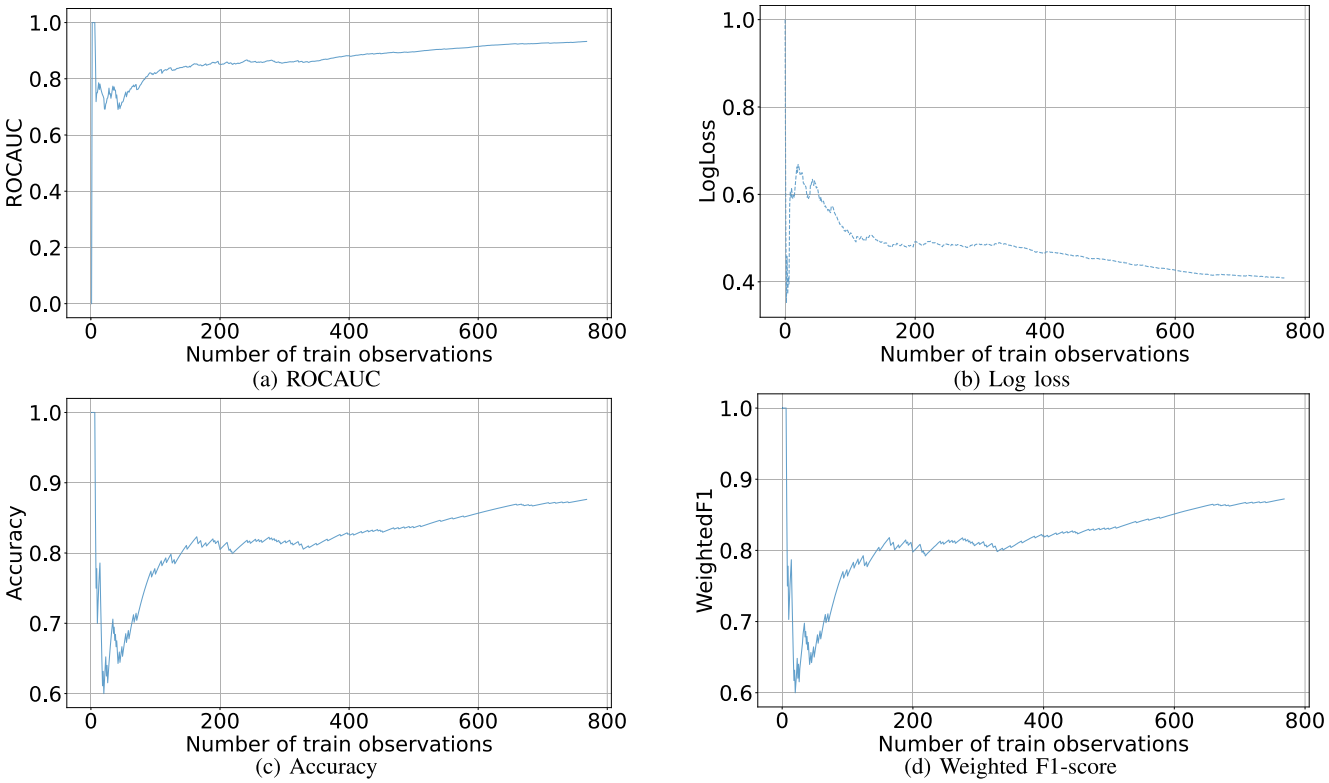


FIGURE 3. **Plots of ROCAUC, log loss, accuracy and weighted F1-score of the proposed incremental algorithm for clasfication.**

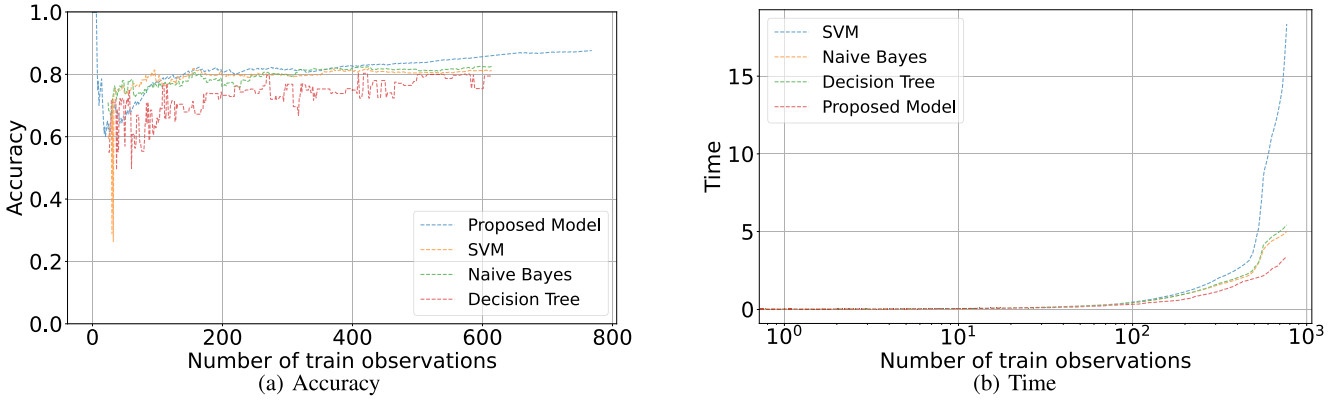


FIGURE 4. **Comparison of the proposed incremental algorithm with the batch learning algorithms for classification.**

TABLE 13. **Performance comparison of the incremental models for anomaly detection.**

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression(baseline)	90.92	82.66	93.72	87.84
One Class SVM(Quantile Filter) [32]	95.90	92.37	96.23	94.26
One Class SVM(Threshold Filter) [32]	92.68	93.55	84.94	89.04
ALMA [27]	86.97	75.86	92.05	83.18
Proposed model(Quantile Filter)	95.75	92.00	96.23	94.07
Proposed model(Threshold Filter)	**97.22**	**95.45**	**96.65**	**96.05**

The threshold filter used over here classifies the anomaly score based on a fixed threshold value. Any instance having score above the threshold is classified as an anomaly. On the other hand, the quantile filter classifies the anomaly score based on a running quantile of anomaly scores and classifies any score above current quantile as an anomaly. The quantile

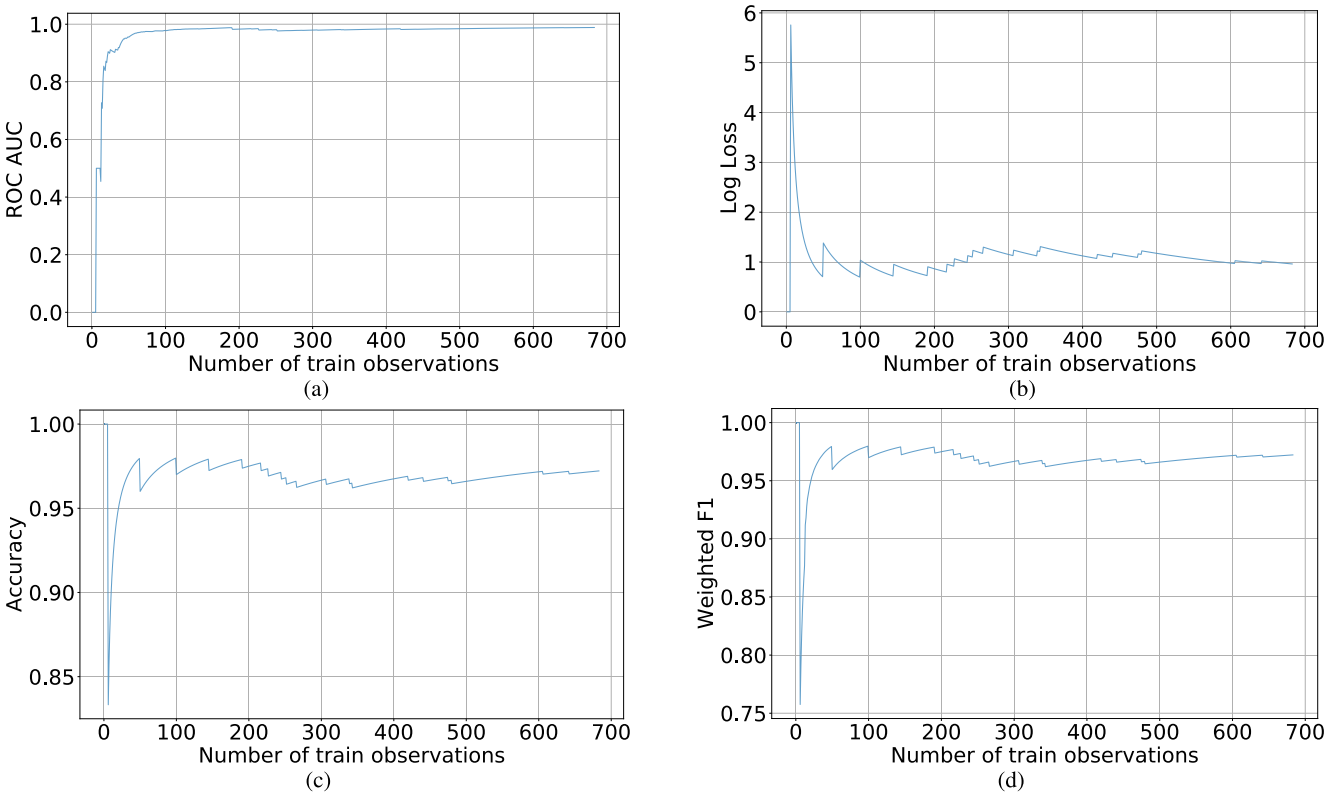


FIGURE 5. Plots of ROCAUC, log loss, accuracy and weighted F1-score of the proposed incremental algorithm for anomaly detection.

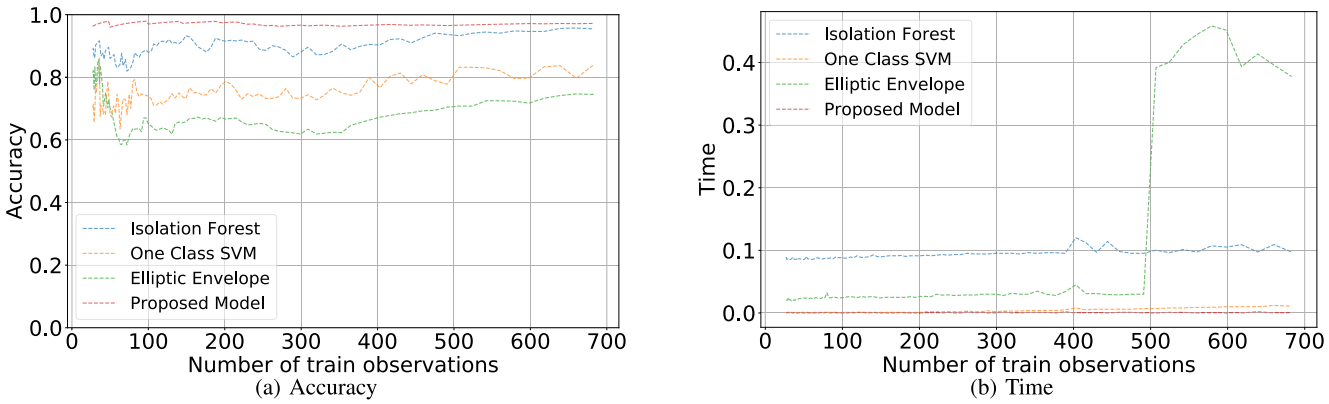


FIGURE 6. Comparison of the proposed incremental algorithm with the batch learning algorithms for anomaly detection.

is updated with each new observed score allowing the filter to adapt to change in data over time.

Table 12 shows the accuracy and training time of different ensemble learning models and planned models. The proposed model outperforms other collective learning models in terms of both accuracy and training time. The accuracy of the model is 97.22%, which is 1.76% higher than the classified forest model. The isolation forest model performed best among all batch-learning models. The OneClassSVM model used here is also a learning-based model, different from the model used above for incremental anomaly detection, which is a stochastic use of a class of SVM algorithms.

Figure 5 presents the plots of ROCAUC, log loss, accuracy and weighted F1-score of the proposed incremental algorithm for classification. Figure 5a shows the ROCAUC score's variation with each dataset point incremented to the model. Overall the ROCAUC score increases with each update in the model until the number of train observations is 70. It reaches its peak value around 70 train observations, after which the ROCAUC score is nearly constant. Figure 5b is a log loss curve. The curve initially gives a high value and then drops down. It fluctuates around log loss equal to 1, and gradually, the fluctuation weakens and starts to settle around 1. The figure 5c, which is the accuracy curve, the curve first dips

at initial train observations. Then it increases gradually until the number of train observations equals 50. After the number of observations equals 50, the accuracy is nearly constant throughout the domain. The weighted F1 figure 5d has a similar trend as the accuracy curve. The curve first dips at initial train observations. This dip is more profound than the one in the accuracy curve. Then till the number of train observations is equal to 50, the weighted f1 increases. After this, the curve fluctuates a bit and slowly becomes constant.

Figure 6 compares the proposed incremental algorithm with the batch learning algorithms for anomaly detection. From Figure 6a, it can be inferred that the proposed model outperforms the other three models in all analyses. The accuracy of the proposed model was nearly the same after every training iteration. Figure 6b compares the time models take to get trained for different numbers of train observations. The proposed model takes the shortest time for almost all big data training.The training data size is equal to the number of train observations here. Among the batch learning algorithms, the elliptic envelope performs the best in terms of the time taken to train. Our proposed model also outperforms the elliptic envelope model, especially in many training observations, as seen in the figure.

V. CONCLUSION

In this paper, we have presented an edge-based IoMT framework with an incremental machine learning approach, incorporating a novel feature selection algorithm based on bijective soft sets, TOPSIS, and Shannon entropy. We proposed two incremental learning algorithms for classification and anomaly detection. Our proposed model for both tasks achieved good results and performed relatively better than other incremental and traditional batch-based learning models. We perform a detailed experimental evaluation comparing the proposed models with the existing incremental learning algorithms and batch learning-based algorithms. We use metrics such as accuracy, precision, recall and F1 score. The Diabetes dataset and the Breast Cancer dataset are used to evaluate the effectiveness of the proposed system, baseline system, and existing supplementary study system. The proposed model for classification gives an accuracy of 87.63%, which is better by 13.61% than the best performing batch learning-based model. Similarly, the proposed model for anomaly detection gives an accuracy of 97.22%, which is better by 1.76% than the best performing batch-based model. The proposed incremental algorithms for classification and anomaly detection are 9X and 16X faster than their corresponding best-performing batch learning-based models.

ACKNOWLEDGMENT

(Kartikey Singh Bhandari and Aditya Nawal contributed equally to this work.)

REFERENCES

[1] S. Vishnu, S. J. Ramson, and R. Jegan, "Internet of Medical Things (IoMT)—An overview," in *Proc. 5th Int. Conf. Devices, Circuits Syst. (ICDCS)*, 2020, pp. 101–104.

[2] (2026). *Iot Medical Devices Market—Global Forecast to 2026 | Marketsandmarkets*. Accessed: Mar. 23, 2023. [Online]. Available: https://www.marketsandmarkets.com/Market-Reports/iot-medical-device-market-15629287.html

[3] (2028). *Iot Healthcare Market Size, Share & Forecast To 2028*. Accessed: Mar. 23, 2023. [Online]. Available: https://www.marketsandmarkets.com/Market-Reports/iot-healthcare-market-160082804.html

[4] S. Sai, V. Chamola, K. R. Choo, B. Sikdar, and J. J. P. C. Rodrigues, "Confluence of blockchain and artificial intelligence technologies for secure and scalable healthcare solutions: A review," *IEEE Internet Things J.*, vol. 10, no. 7, pp. 5873–5897, Apr. 2023.

[5] S. Sai, V. Hassija, V. Chamola, and M. Guizani, "Federated learning and NFT-based privacy-preserving medical-data-sharing scheme for intelligent diagnosis in smart healthcare," *IEEE Internet Things J.*, vol. 11, no. 4, pp. 5568–5577, 2023.

[6] C. Iwendi, S. Khan, J. H. Anajemba, A. K. Bashir, and F. Noor, "Realizing an efficient IoMT-assisted patient diet recommendation system through machine learning model," *IEEE Access*, vol. 8, pp. 28462–28474, 2020.

[7] M. A. Sayeed, S. P. Mohanty, E. Kougianos, and H. P. Zaveri, "Neuro-detect: A machine learning-based fast and accurate seizure detection system in the IoMT," *IEEE Trans. Consum. Electron.*, vol. 65, no. 3, pp. 359–368, Aug. 2019.

[8] M. F. Khan et al., "An IoMT-enabled smart healthcare model to monitor elderly people using machine learning technique," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–10, Nov. 2021.

[9] S. Y. Siddiqui et al., "IoMT cloud-based intelligent prediction of breast cancer stages empowered with deep learning," *IEEE Access*, vol. 9, pp. 146478–146491, 2021.

[10] K. Datta Gupta, D. K. Sharma, S. Ahmed, H. Gupta, D. Gupta, and C.-H. Hsu, "A novel lightweight deep learning-based histopathological image classification model for IoMT," *Neural Process. Lett.*, vol. 55, no. 1, pp. 205–228, Feb. 2023.

[11] A. Gepperth and B. Hammer, "Incremental learning algorithms and applications," in *Proc. Eur. Symp. Artif. Neural Netw. (ESANN)*, 2016, pp. 1–13.

[12] M. A. Khan and F. Algarni, "A healthcare monitoring system for the diagnosis of heart disease in the IoMT cloud environment using MSSO-ANFIS," *IEEE Access*, vol. 8, pp. 122259–122269, 2020.

[13] A. G. Hussien et al., "Crow search algorithm: Theory, recent advances, and applications," *IEEE Access*, vol. 8, pp. 173548–173565, 2020.

[14] C. Peng, Z. RuiYang, and D. ChunHao, "Dynamic hidden variable fuzzy broad neural network based batch process anomaly detection with incremental learning capabilities," *Exp. Syst. Appl.*, vol. 202, Sep. 2022, Art. no. 117390. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417422007357

[15] K. Raghuraman, M. Senthurpandian, M. Shanmugasundaram, and V. Vaidehi, "Online incremental learning algorithm for anomaly detection and prediction in health care," in *Proc. Int. Conf. Recent Trends Inf. Technol.*, Apr. 2014, pp. 1–6.

[16] S. Ozawa, S. Pang, and N. Kasabov, "Incremental learning of chunk data for online pattern classification systems," *IEEE Trans. Neural Netw.*, vol. 19, no. 6, pp. 1061–1074, Jun. 2008.

[17] M. Masana, X. Liu, B. Twardowski, M. Menta, A. D. Bagdanov, and J. van de Weijer, "Class-incremental learning: Survey and performance evaluation on image classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 5, pp. 5513–5533, May 2023.

[18] A. A. Nancy, D. Ravindran, P. M. D. R. Vincent, K. Srinivasan, and D. G. Reina, "IoT-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning," *Electronics*, vol. 11, no. 15, p. 2292, Jul. 2022.

[19] V. Hayyolalam, M. Aloqaily, O. Ozkasap, and M. Guizani, "Edge intelligence for empowering IoT-based healthcare systems," *IEEE Wireless Commun.*, vol. 28, no. 3, pp. 6–14, Jun. 2021.

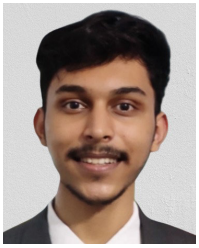
[20] M. Akter, N. Moustafa, T. Lynar, and I. Razzak, "Edge intelligence: Federated learning-based privacy protection framework for smart healthcare systems," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 12, pp. 5805–5816, Dec. 2022.

[21] K. Gong, Z. Xiao, and X. Zhang, "The bijective soft set with its operations," *Comput. Math. Appl.*, vol. 60, no. 8, pp. 2270–2278, Oct. 2010.

[22] T.-C. Wang and H.-D. Lee, "Developing a fuzzy TOPSIS approach based on subjective weights and objective weights," *Exp. Syst. Appl.*, vol. 36, no. 5, pp. 8980–8985, Jul. 2009.

[23] J. Mourtada, S. Gaïffas, and E. Scornet, "AMF: Aggregated Mondrian forests for online learning," 2019, *arXiv:1906.10529*.
[24] B. Lakshminarayanan, D. M. Roy, and Y. W. Teh, "Mondrian forests: Efficient online random forests," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.
[25] S. C. Tan, K. M. Ting, and T. F. Liu, "Fast anomaly detection for streaming data," in *Proc. 22nd Int. Joint Conf. Artif. Intell.*, 2011, pp. 1–6.
[26] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, and Y. Singer, "Online passive aggressive algorithms," School Comput. Sci. Eng., Hebrew Univ., Jerusalem, Israel, Tech. Rep., 2006.
[27] C. Gentile, "A new approximate maximal margin classification algorithm," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 13, 2000, pp. 1–7.
[28] M. Sugiyama, T. Idé, S. Nakajima, and J. Sese, "Semi-supervised local Fisher discriminant analysis for dimensionality reduction," *Mach. Learn.*, vol. 78, nos. 1–2, pp. 35–61, Jan. 2010.
[29] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proc. 8th IEEE Int. Conf. Data Min.*, Dec. 2008, pp. 413–422.
[30] *Diabetes Dataset—Kaggle*. Accessed: Jul. 11, 2023. [Online]. Available: https://www.kaggle.com/datasets/mathchi/diabetes-data-set
[31] W. H. Wolberg and O. L. Mangasarian, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology," *Proc. Nat. Acad. Sci. USA*, vol. 87, no. 23, pp. 9193–9196, Dec. 1990.
[32] K.-L. Li, H.-K. Huang, S.-F. Tian, and W. Xu, "Improving one-class SVM for anomaly detection," in *Proc. Int. Conf. Mach. Learn. Cybern.*, 2003, pp. 3077–3081.

SIVA SAI received the B.E. degree in computer science and the M.Sc. degree (Hons.) in economics from the Birla Institute of Technology and Science, Pilani (BITS Pilani). He is currently a Research Scholar with the Department of Electrical and Electronics Engineering (EEE), BITS Pilani. He worked on multiple research problems, including WiFi CSI activity recognition, multimodal hate speech detection, multilingual offensive/fake speech identification, NLP technologies for lower-resource languages, and deep learning for time series analysis. His past publications were accepted by prestigious journals, magazines, and conferences, such as IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE TRANSACTIONS ON CONSUMER ELECTRONICS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON MACHINE LEARNING IN COMMUNICATIONS AND NETWORKING (TMLCN), IEEE INTERNET OF THINGS, *ACM Transactions on Internet Technology*, *Neural Networks* (Elsevier), EACL, AAAI, *Fire*, EMNLP, *IEEE Consumer Electronics Magazine*, and *IEEE Internet of Things Magazine*. His research interests include applications of blockchain and machine learning for healthcare, natural language processing, computer vision, and connected vehicles.



KARTIKEY SINGH BHANDARI received the B.E. degree in mechanical engineering from the Birla Institute of Technology and Science, Pilani (BITS Pilani), India, in 2022, where he is currently pursuing the master's degree in software systems. His research interests include applications of machine learning in the fields of healthcare, computer vision, and natural language processing.

ADITYA NAWAL received the degree in electrical and electronics engineering from the Birla Institute of Technology and Science, Pilani (BITS Pilani), India, in 2022. He is currently a Software Engineer at Microsoft Corporation. His main research interests include machine learning, artificial intelligence, and security.

VINAY CHAMOLA (Senior Member, IEEE) received the B.E. degree in electrical and electronics engineering and the master's degree in communication engineering from the Birla Institute of Technology and Science, Pilani (BITS Pilani), Pilani, India, in 2010 and 2013, respectively, and the Ph.D. degree in electrical and computer engineering from the National University of Singapore, Singapore, in 2016. In 2015, he was a Visiting Researcher with the Autonomous Networks Research Group (ANRG), University of Southern California, Los Angeles, CA, USA. He was a Post-Doctoral Research Fellow with the National University of Singapore, Singapore. He is currently an Associate Professor with the Department of Electrical and Electronics Engineering, BITS Pilani, where he heads the Internet of Things Research Group/Laboratory. He is also the Co-Founder and the President of a healthcare startup Medsupervision Pvt. Ltd. His research interests include the IoT security, blockchain, UAVs, VANETs, 5G, and healthcare. He is a fellow of IET. He serves as the Co-Chair for various reputed workshops, such as IEEE GLOBECOM Workshop 2021, IEEE INFOCOM 2022 Workshop, IEEE ANTS 2021, and IEEE ICIAfS 2021, to name a few. He serves as an Area Editor for the *Ad Hoc Networks* (Elsevier) and the *IEEE Internet of Things Magazine*. He also serves as an Associate Editor for IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE NETWORKING LETTERS, *IEEE Consumer Electronics Magazine*, *IET Quantum Communications*, *IET Networks*, and several other journals. He is listed among the World's Top 2% Scientists identified by Stanford University.

BIPLAB SIKDAR (Senior Member, IEEE) received the B.Tech. degree in electronics and communication engineering from North Eastern Hill University, Shillong, India, in 1996, the M.Tech. degree in electrical engineering from Indian Institute of Technology Kanpur, Kanpur, India, in 1998, and the Ph.D. degree in electrical engineering from the Rensselaer Polytechnic Institute, Troy, NY, USA, in 2001. He was a Faculty Member with the Rensselaer Polytechnic Institute from 2001 to 2013, where he was an Assistant Professor and an Associate Professor. He is currently a Professor and the Head of the Department of Electrical and Computer Engineering, National University of Singapore, Singapore. His current research interests include wireless networks and security for the Internet of Things and cyber-physical systems. He has served as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE INTERNET OF THINGS JOURNAL, and IEEE OPEN JOURNAL OF VEHICULAR TECHNOLOGY.