International Journal on Applications in Engineering and Technology
Volume 8 : Issue 2 : May 2022, pp 38 – 47 www.aetsjournal.com ISSN (Online) : 2455 - 0523

Disease Prognostication with Concealing Conservancy for Dispensation Structured Concerning Health

PRABHU.S[1], ABIRAMI.G[2], DHARANI.K[3], GAYATHRI.M[4], KEERTHANA.R.S[5]

[1]Assistant Professor, Department of Computer Science and Engineering,
Paavai College of Engineering

[2345]Undergraduate student, Department of Computer Science and Engineering,
Paavai College of Engineering

Abstract: - Artificial intelligence and machine learning have recently gotten a lot of press in the medical field. Machine learning algorithms in healthcare applications frequently incorporate data from a variety of sources, such as hospitals or patients' own devices. One major issue is in evaluating such data without jeopardising the privacy and personal data of patients, which is a significant concern.In healthcare applications, this is a worry. As a result, we're interested in machine execution in these applications.Without revealing sensitive information about the data subjects, learning techniques over scattered data is possible.We present a distributed very randomised trees technique for learning from dispersed data in this paper.data with the goal of maintaining privacy. We present the implementation of our technique (which we refer to as k-PPD-ERT) on a cloud platform and demonstrate its performance using medical data from the Norwegian INTROducing Mental Health through Adaptive Technology (INTROMAT) project, including Heart Disease, Breast Cancer, and mental health datasets (Depresjon and Psykose datasets).

Keywords: Distributed learning , extremely randomized trees, privacy-preserving machine learning, structured health data

1. INTRODUCTION

AI and automated decision-making have the potential to improve accuracy and efficiency. In particular, AI has been shown to outperform humans.create medical expertise in specific fields Here are two examples:Electrocardiography signals are classified according to their rhythms.[1] and breast cancer prediction via deep neural networks cancer utilising the AI system described in [2]; more information [3], [4] are two studies that can be discovered. The application, on the other hand,for automated decision-making using AI and machine learning .There are issues in healthcare, such as security and confidentiality.privacy.For example, if sharing a patient's medical data with a third-party data recipient indicates that he or her has a medical problem, the patient's privacy is breached. As time passes, this becomes more apparent.hard because, in healthcare systems, data may be dispersed across a range of sources rather than being centralised.being kept in a centralised databaseHospitals in scattered settings must utilise data mining techniques to discover useful patterns from patient data.Although hospitals may be able to employ their own resources,restricted resources and health information stored locallydata mining, which is the use of publicly available health data across a number of hospitals leads to the acquisition of more useful and information that is correct However, this is a difficult task because to issues of privacy and legality Hospitals are frequently required to comply.with privacy laws that limit the sharing of health data Sharing regarding patients with third parties, such as other hospitals, is prohibited.Specialists and family doctors [5], [6]. A similar issue exists.when the data is disseminated via the patients' personal networks Previously, it was thought that all sources holding a portion of the data would share their data with a trustworthy third party.However, because the privacy of data sources cannot be secured from a third party, such an assumption, i.e. placing this level of trust in a third party, is not viable in ever circumstance [12]. One method for addressing the privacy issue would be to disturb the data before releasing it. Perturbation-based systems, on the other hand, have limits in terms of data privacy and utility [13], [14]. This is because if the disturbance is not carefully controlled, the data's utility will be reduced, and privacy will be compromised [14].Anonymization approaches, such as [15]– [20], have a similar effect., for example, [15]– [20], share an altered version of data to prevent data subjects from being re-identified [21]. Furthermore, approaches that provide differentiated privacy [22] exchange info while maintaining individual privacy individuals by increasing the amount of noise. Nonetheless, these techniques are not without flaws.Between data privacy and data availability, there is always a trade-off.[13] utilityFor privacy-preserving data mining, previous works have looked at cryptographic algorithms and safe multi-party computation methods [23]– [25]. Such, on the other hand,techniques are ineffective, especially when dealing with large-scale

projects.Due to extensive connectivity and computation, it is difficult to scale data.The price of a tation [14]. Several strategies, such as [12], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37],have been presented to deal with these kinds of costs in machine learning algorithms that preserve privacy, as well as enhance their effectivenessWe address the problem of learning from data from numerous sources without explicitly sharing raw data in this study. The learning data is assumed to be horizon-free.tally partitioned, which means that distinct data records are stored separately.gathered from several sources We concentrate on classification.health data that is organised and can be saved in spreadsheets.We extend our earlier work [28] by proposing a scalable privacy-preserving framework for distributed machine learning based on highly randomised data.trees algorithm, which has a linear cost in the number of iterations.a number of parties and is capable of handling missing values. We're talking about k-PPD-ERT (Privacy-Preserving Distributed RT) is our technique.Extremely Randomized Trees), where k is the number of nodes in the tree.in our method of collaborating partiesFor performance evaluation, we employ two widely used publically available healthcare datasets: Heart Disease [29] and Breast Cancer [30].Datasets from Wisconsin (Diagnostic) [30]. This information represents where there are missing values in medical applications, and Our algorithm is built to deal with situations like this. Finally,On Amazon's website, we demonstrate the application of our technique.AWS cloud and put it to the test in a real-world scenario. on the mental health databases linked to the Norweigian .INTRODUCING Adaptive Technology to Improve Mental Health [31] ogy (INTROMAT) projec.tThe rest of this paper is laid out as follows.Section II examines the current state of distributed privacy-preserving machine learning algorithms in order to address the problem.the subject of discussion The background is covered in Section III.connected to the technique for severely randomised trees and Multi-party computation that is secure. We show this in Section IV.the k-PPD-ERT approach we proposed, which is an adaptionfor dispersed situations, and an expansion of the ERT method.A simple example is used in Section V to demonstrate the distributed very randomised trees approach. In the sixth section,We assess the system's performance, overhead, and privacy.approach that has been offered Section VII is the final section of the report.this document

1.1 THE CURRENT STATE OF THE ART

The concept of collaborative learning from distant data has gotten a lot of attention recently.In the paper, a set of distributed learning approaches is proposed.literature that does not expressly address the issue of privacy [26],[32]– [34]. Nonetheless, by restricting the quantity of data collected, such strategies indirectly help to the preservation of privacy.It must be shared with others or transferred to a central location .The cloud or the servers.Several research [35]– [38] have used randomization to protect the anonymity of persons in data.techniques for mining For example, before distributing and performing, a process that integrates noise into raw data.[35] proposes data mining processes. Nonetheless, the Noise removal techniques can be used to approximate original values. Hence, Noise removal techniques can be used to approximate original values. As a result, such techniques do not ensure adequate privacy.[14], [39]– [41] promises.Several investigations have used secure multi-party computation (SMC). to execute data [12], [23]– [25], [42], [43] mining data from various sources, with no single point of failure .Except for the mining results, private information should be made public. We're interested in the result of a computation in SMC, but we don't know the secret values required for it.computation. As a result, SMC-based approaches are commonly used.in the learning process, compute interim results without,letting other people in on the secret Despite the fact that such methods can meet the standards for privacy, and the incorporation of safe compute approaches that are inefficient and homomorphic.The method's encryption can significantly raise communication and processing overheads. As a result, Several studies [23], [24], [44] have used cryptographic approaches to achieve privacy [14]. These methods use various data mining algorithms to solve classification, clustering, anomaly detection, and other issues [45]– [48].However, such solutions typically have high communication and processing overheads, making them unworkable in many situations.Managing enormous amounts of data [49].Federated learning has been offered as a way for people to study together.train a model with one party's orchestration but with decentralised training data [26], [32], [50]. Several comprehensive reviews of the state-of-the-art federated networks.In [51]– [53], machine learning techniques are used. The majority of past research in this field has focused on Algorithms for deep neural networks In such neural network techniques, in addition to the contributions of data-holder parties, i.e.Sharing model parameters, like gradients, is a privacy risk.

1.2 THE BACKGROUND

We give a quick review of the highly randomised trees (ERT) method and secure multi-party computation (SMC) in this section, which serve as the foundation for our work.A framework for networked machine learning that maintains privacy.

THE ERT ALGORITHM (A)

The ERT [77] algorithm is a tree-based ensemble learning algorithm.Due to its popularity, it's been widely employed to solve categorization difficulties.Its learning capabilities and resistance to overfitting Tree-based ensemble learning has a number of qualities.[62], [78], and [79] are algorithms. The typical ERT, on the other hand,When data is saved in a central location, an algorithm is utilised.The ERT technique is adapted for distributed environments in which data is

stored and effectively dispersed between multiple parties. We'll go through some of the benefits of in the following paragraphs.When comparing the ERT algorithm to other current solutions for its application in a distributed environment.First, because the ERT algorithm is based on ensemble learning, it is resistant to overfitting. Weak learners are included in ensemble learning methods to create weak classes sifiers that are not dependent on other classifiers created.As a result, according to Condorcet's jury theorem (1785) [80],this ensemble of learnt classifiers' majority vote forecasts better than an individual classifier's vote,and as the number of classifiers grows, so does the accuracy[81] improves As a result, while using the ensemble learning method,Instead of just one, we create a collection of classifiers.For instance, in [12], and lastly, based on the voting results of the classifiers that have been learned When using ensemble learning methods, it's important to keep in mind that The learning algorithm's randomization parameters induce generating classifiers that are distinct from one anotherThe randomization of candidate characteristics and the splitting point for every decision node in the tree are the randomness in the ERT algorithm.parameters [77], resulting in the learning of several classifiers.The ERT method is based on the bagging in ensemble logic.learning. By voting, bagging combines the learnt classifiers.i.e., it forecasts based on the learnt majority vote.classifiers. Bagging, while not increasing bias, does result in Because we are averaging, our learnt model has a decreased variance.because the trained model's decreased variance minimises the risk Overfitting is a type of overfitting [78].Second, ERT is built on trees, and tree-based algorithms have been found to outperform other techniques for the type of structured data we're dealing with. The authors report in [62] that for data in a tabular format with each data attribute listed separatelymeaningful and where we don't have a robust multi-scale infrastructure learnt models from structures relating to time or space In most cases, tree-based algorithms outperform models produced via machine learning.[26], [32] are examples of conventional deep neural networks. Moreover,The interpretability of the data in health-care applications is critical.It is preferable to use learnt models. The patterns that tree-based models produceHowever, because ERT is an ensemble learning approach, instead of learning a model with a single tree as in the ID3 algorithm [64], ensemble methods train a model with several trees.As a model, the programme creates many trees. As a result,When compared to other methodologies, such approaches have a lower explainability.The ID3 algorithm is a method of identifying music.

1.3 PRIVACY PRESERVING DISTRIBUTED EXTREMELY RANDOMIZED TREE

Extremely randomized trees for privacy protection.The proposed solution is presented in this part, which is based on the strategy for extremely randomised trees (ERT) and the SMC stands for secure multi-party computation. As a man-We call our method k-PPD-ERT, as mentioned in Section I.In our technique, k is the number of collaborating partie.In the process of secure aggregation It should be noted that k is a parameter.This can be tweaked to meet your privacy needs. The On the one hand, the algorithm protects privacy; on the other hand, the algo-The raw data is not directly shared, and the algorithm is disseminated.The partial information, on the other hand, is aggregated.Using a multi-party computation technique that is secure. Finally,Our proposed approach is based on the ERT (Extremely Risky Technology).Algorithm for Randomized Trees in.

II. SECURED MULTI PARTY COMPUTING

PROTECTED MULTI-PARTY COMPUTING Yao's Millionaires' dilemma [82] inspired the secure multi-party computation architecture, which considers the problem of a group of people working together to solve a problem which has a hidden value Both parties have an interest in the outcome of a calculation based on their top-secret information values, while refusing to reveal their hidden values.

with other individuals. Revealing secret values with a party that everyone trusts is a straightforward way for computing the desired value without sharing secret information with other parties. The trustworthy party has the ability to then run the code and return the result to everyone parties. The assumption of trusted parties, on the other hand, is not correct.Because the privacy of persons with whom you interact is protected in many instances, it's possible. As such, hidden values cannot be safeguarded from third parties. Solutions aren't feasible. As a result, depending on the type of Other than the computation and scenarios, we'll need to come up with something else strategies for completing the specified collaborative computation in a secure manner and without invading one's privacy.We offer a basic approach for secure aggregation of secret values to demonstrate SMC. The mechanism for secure aggregation is depicted in Figure 1. We have four parties in this scenario.Each party is in possession of a secret value (S. V.), and the parties are interconnected. P4 is the result of adding all secret values together.i=1,S.V.i. To securely aggregate the secret values, use the following formula:
(i) The first party creates a random mask, which is then aggregated.with the value of its secret value (S. V.1), and transmits the result to the next get-together

(ii) The input is received by the following parties, who combine it with other data.Send the result to the next party based on their secret values.The outcome is sent to the first party by the last party.

(iii) The result is given to the first party.As a result, based on the information obtained, each party cannot determine the secret value of the prior parties. In this strategy, however, if two adjacent parties, i.e., the par

They cooperate before and after a certain celebration in the ring.will be able to figure out what the victim's secret value is.

For example, if Party 2 reveals Party 3's input,Party 4 releases the output of Party 3 at the same time, allowing them to reveal the secret value of Party 3. As a result, the bare minimum of collaborating parties required to uncover a secret is,In this procedure, the value is two. In addition, in terms of overhead,

In this method, each secure compute operation,One message is sent and received by each side. Consequently,The communication overhead for this strategy is 2n, where n is the number of participants. the number of political parties Secured multi party computation.

III. ADAPTATION OF ERT FOR DISTRIBUTED SETTINGS

ERT modification for distributed settings The complete technique for learning an ensemble of decision trees using the ERT algorithm is presented in this section.The setting was discussed. The algorithm's pseudocode is also available.for the sake of clarity.

3.1 INITIALIZATION AND START OF THE PROCESS

The start of the learning process and its initialization In our distributed learning framework, we have two categories of parties. We have a mediator who orchestrates and mediates.examines the overall learning process as well as a number of data-holders parties who collaborate with one another, as well as the mediator Learn how to create a categorization model. Algorithm 1 and Algorithm 2 are two different algorithms**.**show the procedures and functions pseudo codes for the parties acting as a mediator and data holders, respectively.

3.2 SHARING THE UNEXPECTED SEEDS

To begin this process, all parties agree on a global seed for the random function (Algorithm 1).Algorithm 2, Line 1 and Algorithm 1, Line 1). The world's seed,The mediator and all data holders share this trait.We have two ran- parameters in the ERT algorithm.domness for learning a classifier that isn't very good. First and foremost, we require to choose several features for the candidate at random At each stage of our decision-making process, we use decision nodes.a tree (Algorithm 1, Line 24 Algorithm 2, Line 25).Second, for each attribute in the dataset, a random splitting point.It is necessary to have a candidate decision node (Algorithm 1).Algorithm 2, Line 26, and Algorithm 2, Lines 28– 35,29– 36 lines). The parties who have the data and the mediator

The global random seed (known to all parties) is set in the mediator in Algorithm

1 Mediator 1.

2 • Wait for the data-holder parties to establish a connection.

3. do for I = 1 to M

4 • Build a tree with ti = Build k-PPD-ERT(0, 'None')

5th and final

6. E = t1, t2,..., tM tM tM tM tM tM tM tM tM tM tM tM tM tM

Build k-PPD-ERT(Split ID, Branch) is a function in the Build k-PPD-ERT package.

8 • Use Secret aggregation(Split ID, Branch) to send a message.

a request to the parties who have the data

9 • Wait for the data-holder to send you the results.

parties

10 • Sum = total of all received findings

parties who own the data

• Generate splits() (11), (based on the global seed)

12 if the number of classified records is less than nmin or if the number of classified records is more than nmax

If the categorised documents' labels are the same,

13 send back a leaf label

14 if not

15 • Calculate the score for each split (Information)

16 • Pick the split with the highest point total.

17 • Notify all parties (for Split ID) of the chosen split.

• construct tree T = construct k-PPD-ERT(next)

Split ID, 'T'))))))))))))))))

19 • Create tree F = Create k-PPD-ERT(next)

Split ID, 'F'))))))))))))))))

20 • Create a node with the split you want to use and attach it to it.

T and F subtrees are tree T and tree F, respectively.

return the tree that was created as a result of the operation.

the twenty-first

end of number 22

Generate splits is a function that generates splits ()

24 • Pick D qualities at random: a1,..., aD

25 • Make D splits: s1, s2, s3,..., sD, where si =

Pick rand split(ai)

s1,...,sD are the 26 return splits.

end of 27

Pick rand split is a function that divides a random number into two parts (a)

29. If an is a category variable, then

30. will give you a possible category.

end of the 31st

if an is a number, then

33 .Return a value that falls between the minimum and maximum values.

finish number 34

end of 35 mediator, who will then share them with all parties for future duties. Because everyone uses the same random seed, the global random seed, they all get the same results.

At each phase, there are candidate decision nodes, but no fundamental changes are made .overhead in communication In addition, for secure partial result aggregation, k selected data-holders, as described in Section IV-B. For the random function, each party sends a unique seed. via secure connectivity to other data owners

(**Line 2 of Algorithm 2**) These seeds were chosen at random and are unique. For any pair of data-holder parties, it is both public and private. (b) Begin the learning process for one decision tree.

The privacy-preserving distributed ERT technique is an ensemble learning method, thus we repeat the decision tree learning process M times until we get it right. M decision trees are required (Algorithm 1, Lines 3– 5). The number of trees, M, is a user-adjustable parameter ,to compromise between robustness and overhead. Due to the fact that we learn distinct decision trees every time, In ERT, there is a lot of randomness. Finally, after going through the process a second time ,the process of learning a decision tree We save the trees M times.in the key of E (Algorithm 1, Line 6). For forecasting the future, the E, which is an ensemble of learning trees, will be used.

3.3 THE PROCESS OF LEARNING ONE DECISION TREE

A recursive approach is used to learn a decision tree using the privacypreserving distributed ERT algorithm. The operation is carried out in a top-down manner, beginning at the root and ending at the leaves. The Split ID or ID for the decision node is zero for the root decision node, and Because there hasn't been a previous branch, the Branch input has been set to
'None' is the answer (Algorithm 1, Line 4).

(a) Candidate Decision Node Generation
Extremely randomised data was used to construct each decision tree.The mediator creates the candidate decision nodes from the tree.
(Line 11 of Algorithm 1) The mediator will then make a decision. based on the optimal decision node among the candidates the information obtained from data-holder parties The candidate decision nodes are produced at random, based on a set of criteria. according to the global random seed.

IV. DESCRIPTION OF THE HARDWARE

A power supply (sometimes known as a power supply unit or PSU) is a device or system that provides electrical or other types of energy to one or more output loads. The phrase is most generally used to describe electrical energy sources, with mechanical energy sources being used less frequently and others being used only infrequently.
This is a simple +5V power supply that comes in handy when working with digital circuits. Any electronics store or supermarket will have small, affordable wall transformers with changeable output voltage. These transformers are inexpensive, but their voltage regulation is typically poor, making them unsuitable for digital circuit experimentation unless greater -regulation can be provided in some way. The following circuit is the solution to the problem.

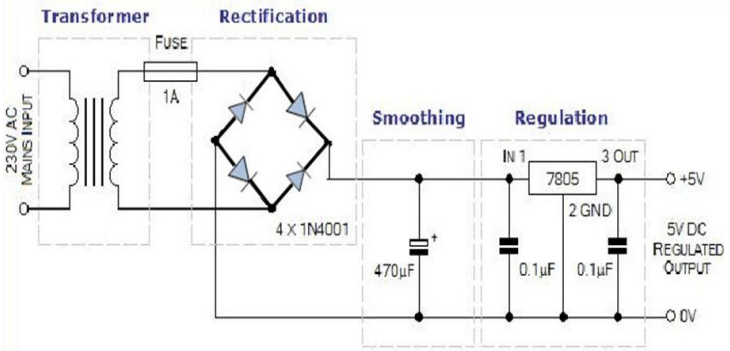

Fig 4.1 : 5volts of Electricity

4.1 THE LM35'S WORKING PRINCIPLE

In the drawing's middle, there are two transistors. The emitter area of one is ten times that of the other. Because the same current flows through both transistors, it has a tenth of the current density. This results in a voltage across resistor R1 that is proportional to the absolute temperature and nearly linear over the whole range. A specific circuit takes care of the "nearly" component, straightening out the slightly curved voltage against temperature graph.

By comparing the outputs of the two transistors, the amplifier at the top assures that the voltage at the base of the left transistor (Q1) is proportional to absolute temperature (PTAT). A continuous current source circuit is the small circle with the I in it. To make a very precise temperature sensor, the two resistors are calibrated in the manufacturing. There are numerous transistors in the integrated circuit, including two in the centre, some in each amplifier, some in the constant current source, and some in the curvature compensation circuit. All of this is included within a three-lead package.

4.2 SENSOR OF TEMPERATURE

A temperature sensor is a gadget that is meant to measure the warmth or coldness of an object. The LM35 is a proportional temperature sensor with a precision IC temperature sensor (in degrees Celsius).The temperature may be measured more precisely with the LM35 than with a thermistor. It also has a low self-heating property, causing a temperature rise of less than 0.1 °C in still air. The temperature range for operation is -55°C to 150°C. The LM35's low output impedance, linear output, and perfect intrinsic calibration make it particularly simple to interface with readout or control circuitry. It has been used in power supplies, battery management, appliances, and other areas.

V. DESCRIPTION OF THE SOFTWARE

Android (stylized as android) is a touch-screen mobile operating system developed by Google, based on the Linux kernel and built primarily for smart phones and tablets. On-screen items are controlled via touch gestures that loosely approximate to real-world motions like swiping, tapping, and pinching, as well as a virtual keyboard for text input, in Android's user interface. Google has also developed Android TV for televisions, Android Auto for cars, and Android Wear for wrist watches, all of which have their own unique user interfaces. Android is also seen on laptops, game consoles, digital cameras, and other electrical devices. The world is becoming smaller as mobile phone technology advances. As the number of customers grows, so do the amount of facilities available. Mobile phones have changed and become a part of our life, starting with simple ordinary handsets that were only used for making phone calls. They are now utilised for more than just making phone conversations, and may be used as a camera, music player, tablet PC, television, and web browser, among other things. New software and operating systems are also required as a result of the new technology.

4.2Android operating system definition: Operating systems have progressed significantly during the previous 15 years. Mobile OS has come a long way, starting with black and white phones and progressing to smart phones and small computers in recent years. Mobile OS has changed significantly in recent years, from Palm OS in 1996 to Windows Pocket PC in 2000, and then to Blackberry OS and Android. Android is a collection of software that includes not only the operating system but ANDROID is currently one of the most widely used smartphone operating systems.also middleware and key applications. Andy Rubin, Rich Miner, Nick Sears, and Chris White created Android Inc in Palo Alto, California, in 2003. Google later purchased Android Inc. in 2005. There have been a number of sequels to the original release.

VI.PRIVACY AND SECURITY

In September 2013, it was revealed that the American and British intelligence agencies, the National Security Agency (NSA) and Government Communications Headquarters (GCHQ), respectively, have access to user data on iPhone, BlackBerry, and Android devices as part of the broader 2013 mass surveillance disclosures. They allegedly have access to practically all smartphone data, including SMS, GPS, emails, and notes. Further reports surfaced in January 2014, revealing the intelligence agencies' ability to intercept personal information sent across the Internet by social networks and other popular applications like Angry Birds, which collect personal information from their users for advertising and other commercial purposes. According to The Guardian, GCHQ has a wiki-style guide to various apps and advertising networks, as well as the various data that may be extracted from each. Later that week, Rovio, the Finnish Angry Birds maker, declared that, in light of these disclosures, it was reviewing its partnerships with its advertising platforms, and urged on the rest of the industry to do the same. According to the documents, spy agencies are attempting to intercept Google Maps searches and queries submitted from Android and other smartphones in order to obtain location data in bulk. Although the NSA and GCHQ assert that its activities are compliant with all applicable domestic and international laws, the Guardian reports that "the latest revelations may add to rising public disquiet about how the technology is used."

VII . CONCLUSIONS

The topic of restricting and summarising diverse data mining techniques employed in the field of medical prediction is tackled in this research. For intelligent and successful diabetic illness prediction via data mining, the focus is on combining diverse algorithms and combinations of several target attributes. Data mining is a useful tool for deriving

meaningful medical rules from medical data, and it plays a significant role in disease prediction and clinical diagnosis. There is a growing interest in utilising categorization to determine whether or not a disease is present. The current study used a large sample of hospitalised patients to demonstrate classification. The classification technique is extremely sensitive to data that is noisy. If there is any noisy data present, it presents major challenges in terms of classification processing capacity. It not only slows down but also impairs the performance of the classification algorithm. As a result, before using a classification technique, all attributes that will subsequently act as noisy attributes must be removed from datasets. We can incorporate preprocessing procedures and classification rule algorithms, such as deep neural network techniques, in this project work for classifying datasets that are uploaded by users. The deep learning technique has produced superior outcomes than other strategies, according to the trial results.

VIII. WORK IN THE FUTURE

In the future, we will likely increase performance efficiency by employing different data mining techniques and algorithms, as well as shrinking the hardware physical scale form factor.

REFERENCES

[1]. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, ‘ ‘ Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network,’ ’ Nature Med., vol. 25, no. 1, pp. 65– 69, Jan. 2019.

[2]. "International evaluation of an AI system for breast cancer detection," McKinney et al. The journal Nature, vol.577, pp. 89– 94, Jan. 2020.

[3]. X. Liu, L. Faes, A. U. Kale, S. K. Wagner, D. J. Fu, A. Bruynseels, T. Mahendiran, G. Moraes, M. Shamdas, C. Kern, J. R. Ledsam, M. K. Schmid, K. Balaskas, E. J. Topol, L. M. Bachmann, P. A. Keane, and A. K. Denniston, "A thorough evaluation and meta-analysis of deep learning performance against health-care professionals in diagnosing disorders from medical imaging," Digitization of the Lancet.. Health, vol. 1, no. 6, pp. e271– e297, Oct. 2019.

[4]. R. Aggarwal, V. Sounderajah, G. Martin, D. S. W. Ting, A. Karthikesalingam, D. King, H. Ashrafian, and A. Darzi, ‘ ‘ Diagnostic accuracy of deep learning in medical imaging: A systematic review and meta-analysis,’ ’ npj Digit. Med., vol. 4, no. 1, p. 65, Dec. 2021.

[5]. S. D. Lustgarten, Y. L. Garrison, M. T. Sinnard, and A. W. Flynn, ‘ ‘ Digital privacy in mental healthcare: Current issues and recommendations for technoJHLJKIU.,;HYTFU logy use,’ ’ Current Opinion Psychol., vol. 36, pp. 25– 31, Dec. 2020.

[6]. D. Pascual, A. Amirshahi, A. Aminifar, D. Atienza, P. Ryvlin, and R. Wattenhofer, ‘ ‘ EpilepsyGAN: Synthetic epileptic brain activities with privacy preservation,’ ’ IEEE Trans. Biomed. Eng., vol. 68, no. 8, pp. 2435– 2446, Aug. 2021.

[7]. A. Saeed, F. D. Salim, T. Ozcelebi, and J. Lukkien, ‘ ‘ Federated selfsupervised learning of multisensor representations for embedded intelligence,’ ’ IEEE Internet Things J., vol. 8, no. 2, pp. 1030– 1040, Jan. 2021.

[8]. F. Forooghifar, A. Aminifar , and D. Atienza, ‘ ‘ Resource-aware distributed epilepsy monitoring using self-awareness from edge to cloud,’ ’ IEEE Trans. Biomed. Circuits Syst., vol. 13, no. 6, pp. 1338– 1350, Dec. 2019.

[9]. D. Sopic, A. Aminifar, A ction on wearable systems,’ ’ IEEE Trans. Biomed. Circuits Syst., vol. 12, no. 5, pp. 982– 992, Oct. 2018.

[10]. D. Sopic, A. Aminifar, A. Aminifar, and D. Atienza, ‘ ‘ Real-time classification technique for early detection and prevention of myocardial infarction on wearable devices,’ ’ in Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS), Oct. 2017, pp. 1– 4.

[11]. R. Zanetti, A. Arza, A. Aminifar and D. Atienza, ‘ ‘ Real-time EEG-based cognitive workload monitoring on wearable devices,’ ’ IEEE Trans. Biomed. Eng., vol. 69, no. 1, pp. 265– 277, Jan. 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9464276, doi: 10.1109/TBME.2021.3092206.

[12]. F. Emekci, O. D. Sahin, D. Agrawal, and A. El Abbadi, ‘ ‘ Privacy preserving decision tree learning over multiple parties,’ ’ Data Knowl. Eng., vol. 63, no. 2, pp. 348– 361, Nov. 2007.

[13]. J. S. Davis and O. "Improving privacy preservation policy in the new information age," according to Osoba. Health Technol., vol.9, no. 1, pp. 65– 75, Jan. 2019.

[14]. J. Vaidya, B. Shafiq, W. Fan, D. Mehmood, and D. Lorenzi describes his concept as "a random decision tree framework for privacy-preserving data mining."IEEE Transactions on. Dependable Secure Comput., vol. 11, no. 5, pp. 399– 411, Sep. 2014.

[15]. P. Jurczyk and L. Xiong, "Distributed anonymization: Achieving privacy for both data subjects and data providers," in Proceedings of the National Academy of Sciences, vol.IFIP Annu. Conf. Data Appl. Secur. Privacy. Berlin, Germany: Springer, 2009. [Online]. Available:https://link.springer.com/chapter/10.1007/978-3-642-03007-9_13

[16]. L. Sweeney, ‘ ‘ K-anonymity: A model for protecting privacy,’ ’ Int. J. Uncertainty, Fuzziness Knowl.-Based Syst., vol. 10, no. 5, pp. 557– 570, Oct. 2002.

[17]. A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, ‘ ‘ Ldiversity: Privacy beyond K-anonymity,’ ’ ACM Trans. Knowl. Discovery Data, vol. 1, no. 1, p. 3, 2007.

[18]. N. Li, T. Li, and S. "T-closeness: Privacy beyond K-anonymity and L-diversity," Venkatasubramanian, in Proc.IEEE 23rd Int. Conf. Data Eng., Apr. 2007, pp. 106– 115.

[19]. A. Aminifar, Y. Lamo, K. Pun, and F. Rabbi, ‘ ‘ A practical methodology for anonymization of structured health data,’ ’ in Proc. 17th Scandin. Conf. Health Informat., 2019, pp. 127– 133. [Online]. Available: https://ep.liu.se/en/conference-article.aspx?series=ecp&issue=161&Article_No=22

[20]. A. Aminifar, F. Rabbi, V. K. I. Pun, and Y. Lamo, ‘ ‘ Diversity-aware anonymization for structured health data,’ ’ in Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Nov. 2021, pp. 2148– 2154.

[21]. [21] Health Informatics—Pseudonymization, International Organization for Standardization, Geneva, Switzerland, Standard ISO 25237:2017, Jan. 2017. [Online].Available:https://www.iso.org/standard/63553.html

[22]. C. Dwork, ‘ ‘ Differential privacy,’ ’ in Proc. 33rd Int. Colloq. Automata, Lang., Program. (ICALP) (Lecture Notes in Computer Science). Berlin, Germany: Springer-Verlag,

2006.[Online].Available:https://link.springer.com/chapter/10.1007/11787006_1

[23]. In Privacy-Preserving Data Mining, M.Kantarcioglu presents "A overview of privacy-preserving algorithms across horizontally partitioned data."Boston, MA, USA: Springer, 2008, pp. 313– 335. [Online]. Available: https://link.springer.com/chapter/10.1007/978-0-387-70992-5_13

[24]. J. Vaidya, ‘ ‘ A survey of privacy-preserving methods across vertically partitioned data,’ ’ in Privacy-Preserving Data Mining. Boston, MA, USA: Springer, 2008, pp. 337– 358.[Online],Available:https://link.springer.com/chapter/10.1007/978-0-387-70992-5_14

[25]. W. Du and 'Building decision tree classifier on private data,' by Z. Zhan, in Proc.IEEE Int. Conf. Privacy, Secur. Data Mining (CRPIT), vol. 14. Australia: Austral. Comput. Soc., 2002, pp. 1–8. [Online]. Available: https://dl.acm.org/doi/10.5555/850782.850784

[26]. J. Konečný, H. Brendan McMahan, D. Ramage, and P. Richtárik, ‘‘Federated optimization: Distributed machine learning for on-device intelligence,’’ 2016, arXiv:1610.02527.

[27]. H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and "Communication-efficient learning of deep networks from decentralised data," Agüera y Arcas, arXiv:1602.05629, 2016.

[28]. A. Aminifar, F. Rabbi, and Y. Lamo, ‘ ‘ Scalable privacy-preserving distributed extremely randomized trees for structured data with multiple colluding parties,’ ’ in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Jun. 2021, pp. 2655– 2659.

[29]. R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J.-J. Schmid, S. Sandhu, K. H. Guppy, S. Lee, and V. Froelicher, ‘ ‘ International application of a new probability algorithm for the diagnosis of coronary artery disease,’ ’ Amer. J. Cardiol., vol. 64, no. 5, pp. 304– 310, Aug. 1989.

[30]. O. L. Mangasarian, W. N. Street, and W. H. Wolberg, ‘ ‘ Breast cancer diagnosis and prognosis via linear programming,’ ’ Oper. Res., vol. 43, no. 4, pp. 570– 577, Aug. 1995.

[31]. INTROMAT (Introducing Personalized Treatment of Mental Health Problems Using Adaptive Technology). Accessed: Dec. 9, 2021. [Online]. Available: https://intromat.no

[32]. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Y. Arcas, ‘ ‘ Communication-efficient learning of deep networks from decentralized data,’ ’ in Proc. 20th Int. Conf. Artif. Intell. Statist., Mach. Learn. Res., PMLR, 2017. [Online]. Available: https://proceedings.mlr. press/v54/mcmahan17a.htm

[33]. V. Smith, S. Forte, C. Ma, M. Takáč, M. Jordan, and M. Jaggi, ‘‘CoCoA: A general framework for communication-efficient distributed optimization,’’ J. Mach. Learn. Res., vol. 18, p. 230, Apr. 2017.

[34]. S. Baghersalimi, T. Teijeiro, D. Atienza, and A. Aminifar, ‘ ‘ Personalized real-time federated learning for epileptic seizure detection,’ ’ IEEE J. Biomed. Health Informat., early access, Jul. 9, 2021, doi: 10.1109/JBHI.2021.3096127.

[35]. R. Agrawal and R. Srikant, ‘ ‘ Privacy-preserving data mining,’ ’ in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2000, pp. 439– 450.

[36]. S. Agrawal and J. R. Haritsa, ‘ ‘ A framework for high-accuracy privacypreserving mining,’ ’ in Proc. 21st Int. Conf. Data Eng. (ICDE), Apr. 2005, pp. 193– 204

[37]. A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, ‘ ‘ Privacy preserving mining of association rules,’ ’ Inf. Syst., vol. 29, no. 4, pp. 343– 364, Jun. 2004.

[38]. S. Rizvi and J. Haritsa, ‘ ‘ Maintaining data privacy in association rule mining,’ ’ in Proc. 28th Int. Conf. Very Large Databases (VLDB). Amsterdam, The Netherlands: Elsevier, 2002, pp. 682– 693.

[39]. H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar, ‘ ‘ On the privacy preserving properties of random data perturbation techniques,’ ’ in Proc. 3rd IEEE Int. Conf. Data Mining, Nov. 2003, pp. 99– 106.

[40]. Z. Huang, W. Du, and B. Chen, ‘ ‘ Deriving private information from randomized data,’ ’ in Proc. ACM SIGMOD Int. Conf. Manage. Data (SIGMOD), 2005, pp. 37– 48.

[41].] M. Kantarcioglu and J. Vaidya, ‘ ‘ An architecture for privacy-preserving mining of client information,’ ’ in Proc. IEEE Int. Conf. Privacy, Secur. Data Mining, vol. 14, Dec. 2002, pp. 37– 42.

[42]. Y. Lindell and B. Pinkas, ‘ ‘ Privacy preserving data mining,’ ’ J. Cryptol., vol. 15, no. 3, pp. 177– 206, 2002.

[43]. C. Clifton, M. Kantarcioglu, J. Vaidya, X. Lin, and M. Y. Zhu, ‘ ‘ Tools for privacy preserving distributed data mining,’ ’ ACM SIGKDD Explor. Newslett., vol. 4, no. 2, pp. 28– 34, Dec. 2002.

[44]. L. T. Phong, Y. Aono, T. Hayashi, L. Wang, and S. Moriai, ‘ ‘ Privacypreserving deep learning via additively homomorphic encryption,’ ’ IEEE Trans. Inf. Forensics Security, vol. 13, no. 5, pp. 1333– 1345, May 2018.

[45]. J. Vaidya and C. Clifton, ‘ ‘ Privacy-preserving outlier detection,’ ’ in Proc. 4th IEEE Int. Conf. Data Mining (ICDM), Nov. 2004, pp. 233– 240.

[46]. G. Jagannathan and R. N. Wright, ‘ ‘ Privacy-preserving distributed Kmeans clustering over arbitrarily partitioned data,’ ’ in Proc. 11th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), 2005, pp. 593– 599.

[47]. X. Lin, C. Clifton, and M. Zhu, ‘ ‘ Privacy-preserving clustering with distributed EM mixture modeling,’ ’ Knowl. Inf. Syst., vol. 8, no. 1, pp. 68– 81, Jul. 2005.

[48]. H. Yu, X. Jiang, and J. Vaidya, ‘ ‘ Privacy-preserving SVM using nonlinear kernels on horizontally partitioned data,’’ in Proc. ACM Symp. Appl. Comput. (SAC), 2006, pp. 603–610.

[49]. B. Pinkas, ‘‘Cryptographic techniques for privacy-preserving data mining,’’ ACM SIGKDD Explor. Newslett., vol. 4, no. 2, pp. 12–19, Dec. 2002.

[50]. J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, ‘ ‘ Federated learning: Strategies for improving communication efficiency,’ ’ 2016, arXiv:1610.05492.

[51]. P. Kairouz et al., ‘ ‘ Advances and open problems in federated learning,’ ’ 2019, arXiv:1912.04977.

[52]. Q. Li, Z. Wen, Z. Wu, S. Hu, N. Wang, Y. Li, X. Liu, and B. He, ‘ ‘ A survey on federated learning systems: Vision, hype and reality for data privacy and protection,’ ’ 2019, arXiv:1907.09693.

[53]. S. Lo, Q. Lu, C. Wang, H. Paik, and L. Zhu, ‘ ‘ A systematic literature review on federated machine learning: From a software engineering perspective,’ ’ ACM Comput. Surv., vol. 54, no. 5, pp. 1– 39, 2021.

[54]. M. Nasr, R. Shokri, and A. Houmansadr, ‘ ‘ Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning,’ ’ in Proc. IEEE Symp. Secur. Privacy (SP), May 2019, pp. 739– 753.

[55]. S. Truex, L. Liu, M. E. Gursoy, L. Yu, and W. Wei, ‘ ‘ Demystifying membership inference attacks in machine learning as a service,’ ’ IEEE Trans. Services Comput., vol. 14, no. 6, pp. 2073– 2089, Nov. 2021.

[56]. K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. Q. S. Quek, and H. V. Poor, ‘ ‘ Federated learning with differential privacy: Algorithms and performance analysis,’ ’ IEEE Trans. Inf. Forensics Security, vol. 15, pp. 3454– 3469, 2020.

[57]. R. Hu, Y. Guo, H. Li, Q. Pei, and Y. Gong, ‘ ‘ Personalized federated learning with differential privacy,’ ’ IEEE Internet Things J., vol. 7, no. 10, pp. 9530– 9539, Oct. 2020.

[58]. R. C. Geyer, T. Klein, and M. Nabi, ‘ ‘ Differentially private federated learning: A client level perspective,’ ’ 2017, arXiv:1712.07557.

[59]. C. Zhuang, T. She, A. Andonian, M. Sobol Mark, and D. Yamins, ‘ ‘ Unsupervised learning from video with deep neural embeddings,’ ’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 9563– 9572.

[60]. H. Purwins, B. Li, T. Virtanen, J. Schluter, S.-Y. Chang, and T. Sainath, ‘ ‘ Deep learning for audio signal processing,’ ’ IEEE J. Sel. Topics Signal Process., vol. 13, no. 2, pp. 206– 219, May 2019.

[61]. A. M. Vartouni, M. Shokri, and M. Teshnehlab, ‘ ‘ Auto-threshold deep SVDD for anomaly-based web application firewall,’ ’ TechRxiv, 2021. [Online]. Available: https://www.techrxiv.org/articles/preprint/AutoThreshold_Deep_SVDD_for_Anomaly-based_Web_Application_ Firewall/15135468, doi: 10.36227/techrxiv.15135468.v1.

[62]. S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, ‘ ‘ From local explanations to global understanding with explainable AI for trees,’ ’ Nature Mach. Intell., vol. 2, no. 1, pp. 56– 67, Jan. 2020.

[63]. W. Fan, H. Wang, P. S. Yu, and S. Ma, ‘ ‘ Is random model better? On its accuracy and efficiency,’ ’ in Proc. 3rd IEEE Int. Conf. Data Mining, Nov. 2003, pp. 51– 58.

[64]. J. R. Quinlan, ‘ ‘ InduMach. Learn., vol. 1, no. 1, pp. 81– 106, Mar. 1986.

[65]. W. Du and Z. Zhan, ‘ ‘ Using randomized response techniques for privacypreserving data mining,’ ’ in Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), 2003, pp. 505– 510.

[66]. K. Cheng, T. Fan, Y. Jin, Y. Liu, T. Chen, D. Papadopoulos, and Q. Yang, ‘ ‘ SecureBoost: A lossless federated learning framework,’ ’ IEEE Intell. Syst., vol. 36, no. 6, pp. 87– 98, Nov. 2021.

[67]. Y. Liu, Z. Ma, X. Liu, S. Ma, S. Nepal, R. H. Deng, and K. Ren, ‘ ‘ Boosting privately: Federated extreme gradient boosting for mobile crowdsensing,’ ’ in Proc. IEEE 40th Int. Conf. Distrib. Comput. Syst. (ICDCS), Nov. 2020, pp. 1– 11.

[68]. L. Zhao, L. Ni, S. Hu, Y. Chen, P. Zhou, F. Xiao, and L. Wu, ‘ ‘ InPrivate digging: Enabling tree-based distributed data mining with differential privacy,’ ’ in Proc. IEEE Conf. Comput. Commun. (INFOCOM), Apr. 2018, pp. 2087– 2095.

[69]. Q. Li, Z. Wen, and B. He, ‘ ‘ Practical federated gradient boosting decision trees,’ ’ in Proc. AAAI Conf. Artif. Intell., 2020, pp. 4642– 4649.

[70]. J. Friedman, ‘ ‘ Greedy function approximation: A gradient boosting machine,’ ’ Ann. Statist., vol. 29, no. 5, pp. 1189– 1232, 2001.

[71]. Y. Liu, Y. Liu, Z. Liu, Y. Liang, C. Meng, J. Zhang, and Y. Zheng, ‘ ‘ Federated forest,’ ’ IEEE Trans. Big Data, early access, May 7, 2020, doi: 10.1109/TBDATA.2020.2992755.

[72]. S. Truex, N. Baracaldo, A. Anwar, T. Steinke, H. Ludwig, R. Zhang, and Y. Zhou, ‘ ‘ A hybrid approach to privacy-preserving federated learning,’ ’ in Proc. 12th ACM Workshop Artif. Intell. Secur. (AISec), 2019, pp. 1– 11.

[73]. T. Kam Ho, ‘ ‘ Random decision forests,’ ’ in Proc. 3rd Int. Conf. Document Anal. Recognit., Aug. 1995, pp. 278– 282.

[74]. L. Breiman, ‘ ‘ Random forests,’ ’ Mach. Learn., vol. 45, pp. 5– 32, Oct. 2001.

[75]. A. Shamir, ‘ ‘ How to share a secret,’ ’ Commun. ACM, vol. 22, no. 11, pp. 612– 613, Nov. 1979.

[76]. A. Aminifar, F. Rabbi, K. I. Pun, and Y. Lamo, ‘ ‘ Privacy preserving distributed extremely randomized trees,’ ’ in Proc. 36th Annu. ACM Symp. Appl. Comput., Mar. 2021, pp. 1102– 1105.

[77]. P. Geurts, D. Ernst, and L. Wehenkel, ‘ ‘ Extremely randomized trees,’ ’ Mach. Learn., vol. 63, no. 1, pp. 3– 42, Apr. 2006.

[78]. J. Friedman, T. Hastie, and R. Tibshirani, The Elements of Statistical Learning (Springer Series in Statistics). New York, NY, USA: Springer, 2001.

[79]. Z. Lipton, ‘ ‘ The mythos of model interpretability,’ ’ Queue, 2018.

[80]. N. D. Condorcet, Essai Sur l’ Application de l’ Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[81]. L. Rokach, Pattern Classification Using Ensemble Methods. Singapore: World Scientific, 2010.

[82]. A. C.-C. Yao, ‘ ‘ How to generate and exchange secrets,’ ’ in Proc. 27th Annu. Symp. Found. Comput. Sci., Oct. 1986, pp. 162– 167.

[83]. E. Bertino, D. Lin, and W. Jiang, ‘ ‘ A survey of quantification of privacy preserving data mining algorithms,’ ’ in Privacy-Preserving Data Mining. Boston, MA, USA: Springer, 2008, pp. 183– 205. [Online]. Available: https://link.springer.com/chapter/10.1007/978-0-387-70992-5_8

[84]. D. Dua and C. Graff, ‘ ‘ UCI machine learning repository,’ ’ School Inf. Comput. Sci., Univ. California, Irvine, CA, USA, 2019. [Online]. Available: http://archive.ics.uci.edu/ml and https://archive.ics.uci.edu/ml/ citation_policy.html

[85]. E. Garcia-Ceja, M. Riegler, P. Jakobsen, J. Tørresen, T. Nordgreen, K. Oedegaard, and O. Fasmer, ‘ ‘ Depresjon: A motor activity database of depression episodes in unipolar and bipolar patients,’ ’ in Proc. 9th ACM Multimedia Syst. Conf., 2018, pp. 472– 477.

ISSN (Online) : 2455 - 0523

BIOGRAPHIES

PRABHU S is an Assistant professor, department of computer science and engineering in Paavai College of Engineering.

ABIRAMI G is an undergraduate student, department of computer science and engineering in Paavai College of Engineering.

DHARANI K is an undergraduate student, department of computer science and engineering in Paavai College of Engineering.

GAYATHRI M is an undergraduate student, department of computer science and engineering in Paavai College of Engineering.

KEERTHANA R S is an undergraduate student, department of computer science and engineering in Paavai College of Engineering.