Journal of Theoretical and Applied Information Technology
15th November 2021. Vol.99. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

THE CLASSIFICATION OF POSSIBLE CORONAVIRUS TREATMENTS ON A SINGLE HUMAN CELL USING DEEP LEARNING AND MACHINE LEARNING APPROACHES

NOUR ELDEEN KHALIFA [1*], GUNASEKARAN MANOGARAN [2,3]
MOHAMED HAMED N. TAHA[1], MOHAMED LOEY [4]

[1]Faculty of Computers & Artificial Intelligence, Cairo University, Cairo 12613, Egypt
[2]University of California, Davis, USA
[3]College of Information and Electrical Engineering, Asia University, Taiwan
[4]Faculty of Computers Artificial Intelligence, Benha University, Benha 13511, Egypt
* Correspondence author

E-mail: [1]nourmahmoud@cu.edu.eg, [2,3]gmanogaran@ieee.org, [1]mnasrtaha@cu.edu.eg, [4]mloey@fci.bu.edu.eg

ABSTRACT

Every major healthcare system is now under the throes of the Coronavirus disease outbreak as it is operating at its maximum capacity. There is an absolute need to establish an appropriate cure for this virus as quickly and efficiently as possible. Advances in deep learning models may play a critical role in SARS-2 discovery by locating a possible treatment. This article's objective is to demonstrate the machine learning and deep learning models approaches for classifying prospective coronavirus treatment on a single human cell. A partial dataset of RXRX.ai which is a publicly available dataset is used in this research. This work targeted to implement a strategy for automatically identifying a single human cell depending on the type of treatment and its concentration level. Throughout this study, we present a DCNN model along with an image processing approach. The systematic approach comprises translating the original dataset's numerical attributes to the image domain, and then incorporating them into DCNN model. In comparison to standard machine learning techniques including such Ensemble, Decision Tree and Support Vector Machine, the experimental findings indicate that the suggested DCNN model for treatment classification (32 categories) obtained a testing accuracy of 98.05 percent. The (Ensemble) algorithm achieves 98.5 percent for the accuracy test in treatment concentration level prognosis, whereas the suggested DCNN model reached 98.2 percent. The classification of treatments and assessing their concentration levels are considerably accurate due to the performance indicators obtained from the experiments.

Keywords: *Deep Learning; Machine Learning; Coronavirus, COVID-19*

1. INTRODUCTION

There was much concern that the SARS virus had spread towards the end of February 2003, throughout the world due to its alarming levels of infections in Asia and outbreaks in the Middle East, and in countries like Russia and Saudi Arabia that had never seen before [1], [2]. This caused people to start sounding the alarm about viruses as they become significant threats in the 21st century. The (WHO) World Health Organization assigned the coronavirus of 2019 the designation 2019-nCov (COVID-19) [3], [4]. The International Committee on Virus Taxonomy (ICTV) recognized the 2019 coronavirus as SARS-CoV-2 in 2020 [5]–[7]. Prior to the publication of this article, an SARS-CoV-2 outbreak claimed over 500 thousand lives in 213 countries and territories [8]. Coronavirus transmission (individual to individual) was spreading at a breakneck pace such as Italy [9], the US [10], India [11], and Germany [12]. By 10 July 2020, almost 12 million individuals associated with SARS had been reported, over 6 million had been recovered, and over 55,000 people had committed

suicide. Figure 1 illustrates some statistics regarding survived and fatal COVID-19 patients [13].

Some of the research's dedicated to different problems related to COVID -19 and solved by computer science field such as tracking of COVID -19 geographical infections on real-time tweets [14], studying the role of emerging technologies for combating COVID -19 pandemic [15], exploring the impacts of COVID -19 on oil and electricity industry [16] and more. The majority of papers are devoted to categorizing and characterization of COVID-19 CT and X-ray images [17]–[20]. The purpose of this research is to evaluate and characterize a medicine capable of assisting in COVID-19 recovery, as well as to explore COVID-19's morphological impact. Today, deep learning is rapidly establishing itself as a fundamental approach for segmenting, classifying and detecting images and videos [21]–[24]. This article will explore the categorization of prospective coronavirus therapies on a single human cell using a machine learning and deep learning model approaches. The goal of this study is to automatically classify a single human cell based on the type of treatment and the concentration of the therapy. The research is novel in that it employs a hypothesized classification model for COVID-19 viral therapies that is dependent on deep learning and machine learning approaches. The rest of the document follows a consistent format. Part 2 summarizes the properties of the data set. Part 3 details the suggested prototype. Part 4 records and evaluates preliminary results, whereas part 5 presents the hypotheses and planned future study.

2. MATERIAL AND METHODS

The experiments in this research were based on the dataset presented in the study [18]. Table 1 contains a detailed concise description of the dataset's properties. The data are freely accessible on RxRx.ai under the title RxRx19a Dataset.

Table 1. Attributes of the RxRx19a Dataset

Attribute	Description
site_id	A site's unique identification
well_id	A given well's unique identification
cell_type	Cell type-determined
experiment	Experiment indicator
plate	The experiment's plate numbering system.
well	Placement on the plate.
site	The location of the image in the well is marked (1, 2, 3, or 4)
disease_condition	The illness state examined in the well (irradiated, viral, or mock).
treatment/therapy	The mixture was evaluated in the well.
treatment-conc/therapy-conc	Concentration of compound evaluated with a measuring unit of in uM.
feature 1 to 1024	Characteristic of the cell (Feature cells have 1024 characteristics.)

It is a multidimensional dataset comprised of over 300,000 documented studies and in a human cell's paradigm of SARS-CoV-2 disease, an examination of approximately 1,660 FDA-approved medicines. Although the offered data is from an in vitro screening of a particular type of human cell, it is probable that this dataset is relevant to a wide variety of additional human primary cell types.

A data subset is incorporated into the study approaches done in this study. VERO cells, a continuous cell lineage obtained from the African

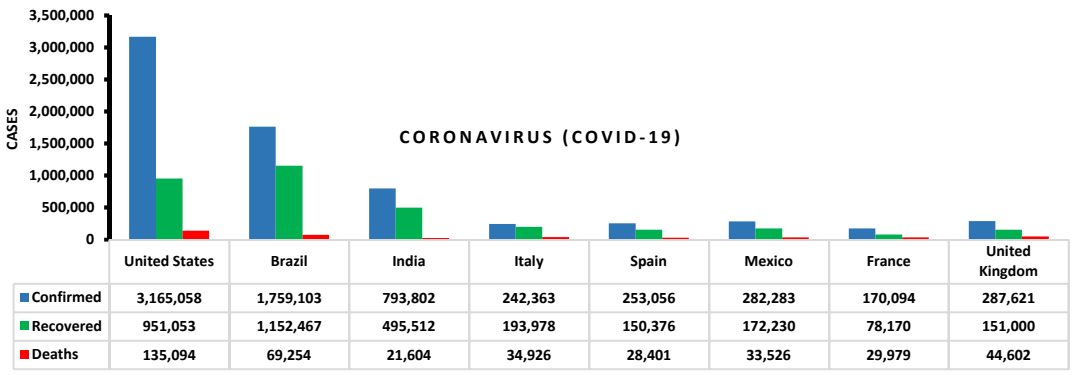

Fig1. Statistics About COVID-19 In Selected Countries

green monkey's kidney epithelial cells, and Human Renal Cortical Epithelial (HRCE) cells comprise the fraction. these cells were chosen in combination with active SARS-CoV-2 at concentrations of 10, 30, and 100f from treatments. This data subset contains 32 medications and three medications concentrations, each of which corresponds to a different type of cell. The experiment conducted in this study includes just 3,750 cell data.

3. THE PROPOSED METHOD AND MODEL

The model presented here is categorized into three stages. The preparation\processing stage changes the 1024 cell numerical values characteristics to a digital image in the first stage. The training second stage, which involves applying machine learning approaches for DCNN and numerical features for transformed image data. The third stage is testing and evaluating the suggested accuracy of the suggested model for treatment classification and prognosis of concentration levels. The suggested approach composition is depicted in Figure 2.

3.1 The Preparation\Processing Stage

The preparation processing stage consists of the following steps: 1) storing the 1024 characteristics of cells in the computer's memory; 2) Using equation (1), transform the original numerical domain of the cell characteristics [-0.00046466477, 4.508815065] to the image region [0, 255]; and 3) By translating the data vector of 1024 capability cells into a 32 x 32-pixel image utilizing the pseudocode described in Algorithm 1, the image is constructed. The output of this stage is going to be 3,750 images. Figure 3 depicts a group of photos following the preparation\processing stage.

$$\text{Value in Pixels} = \text{Round}\left(\frac{(\text{feature cell val} \quad (-0.00046466477))}{4.508815065} * 255\right) \quad (1)$$

-0.00046466477, 4.508815065 and 255 are values indicated to the smallest, biggest numerical data and the biggest data respectively in 1024 characteristics of cell input and in image domain.

Algorithm 1: Building an image from the cell data vector's 1024 features

function build_Image (A)

y is the data vector containing 1024 feature cells as an input.
z represents the resulted image.

```
start
n =1
m=1
for x=1 to 1024 step 1 do
    if x moduls 32 != 0 then
    z[n,m]=y[x];
    n++;
    else
    z[n,m]=y[x];
    m=1;
    n++;
    endif
end for
return z
end
```

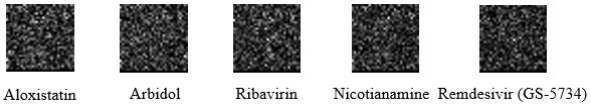

Fig3. *Illustrations Of Transformed Cell Images*

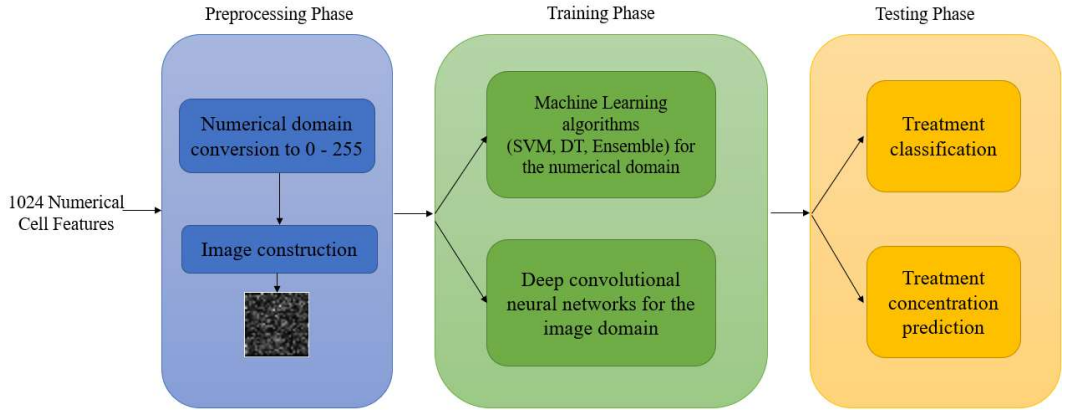

Fig. 2. The Model's Suggested Composition And Phases

3.2 Training Stage

During the training phase, two techniques are used. The initial methodology incorporated machine learning techniques such Ensemble algorithms, Decision Trees, and Support Vector Machines. The second technique makes use of DCNN.

3.2.1 Machine with Support Vectors

SVM is among the best often used and effective machine learning approaches for identification and prediction. As indicated in equation (2), SVM is a functional method, where l is a label between 0 and 1, w.a-k signifies the output, w and q indicate linear class coefficients, and signifies an input. Equation (3) will impose the loss function's reduction [26], [27].

$$SVM_{h_k} = \max(0,1 - l_k(w.a_k - q)) \quad (2)$$

$$SVM_{loss} = \frac{1}{m}\sum_{t=1}^{m} \max(0, h_t) \quad (3)$$

3.2.2 Decision Tree

The decision tree is a categorization paradigm for computers that is centered on the entropy technique and knowledge acquisition. As indicated in Equation 4, entropy calculates the degree of uncertainty in data. Where CD denotes data, b is a category outcome, and p(x) denotes the percentage of q label. We compute knowledge acquisition (KA) by calculating the entropy difference between findings, as given in equation (5), in which x is a data subset [28], [29].

$$\text{Entropy } (CD) = \sum_{i=1}^{n} -\text{p}(\text{ b}_\text{i}\text{ }).\log(\text{ p}(\text{ b}_\text{i}\text{ })\text{ }) \quad (4)$$

3.2.3 Ensemble Methods

Ensemble methods are machine learning algorithms that combine multiple classifiers, which are applied to separate training data to find new cases in either a single class or multiple classes (Typically, this is accomplished via unweighted or weighted voting [30]. Linear Regression is the technique that is employed [31], Logistic Regression [32], and K-Nearest Neighbors Algorithm (k-NN) [33]. We adjust the ensemble using equation (6) to obtain the adequate results [34].

$$\bar{y} = \sum_{k=1}^{h} \alpha_k y_k \quad (6)$$

3.2.4 Deep convolutional Neural Networks (DCNN).

Figure 4 demonstrates the suggested structure of DCNN. The suggested DCNN composes of three primary convolutional layers with a 3*3 pixels window size, three pooling layers and three ReLU layers. The prior levels are required to obtain characteristics, whereas the classification layers consist of two completely connected layers. The current proposal for DCNN is the result of intense refining and tweaking of the architecture based on the work given in [35]–[38].

One issue that DCNN faces is overfitting. Data augmentation can be used to address overfitting [39]–[42]. By applying label-preserving changes to the data, increasing of data enhances the images numbers available applying for training. Additionally, it is implemented to the training set to ensure that the final prototype is more robust to image transformations; in this experiment, each training data image is altered in the following manner:

- Reflection on the X-axis
- Reflection on the X-Y axis.
- Reflection on Y-axis.

The augmentation method increases the collection of images from 3750 to 15000, resulting in a threefold increase in the dataset size. As a result, the training stage of will be improved considerably. Additionally, it will strengthen the suggested DCNN's resistance to data memory and make it stronger.

3.3 Testing Stage

The testing stage demonstrates the suggested model's productivity and effectiveness. The suggested model's primary objectives are to accurately categorize treatments based on numerical characteristic utilizing machine learning methods and classify treatment images dependent on DCNN correctly. Additionally, the treatment concentration for each cell is predicted using both numerical and visual data via machine learning and DCNN.

For machine learning, Accuracy testing and the Receiver Operating Characteristic (ROC) curve following 5kfold Cross-Validation will be included in the performance evaluation. For DCNN assessment, F1 score, the accuracy, precision and

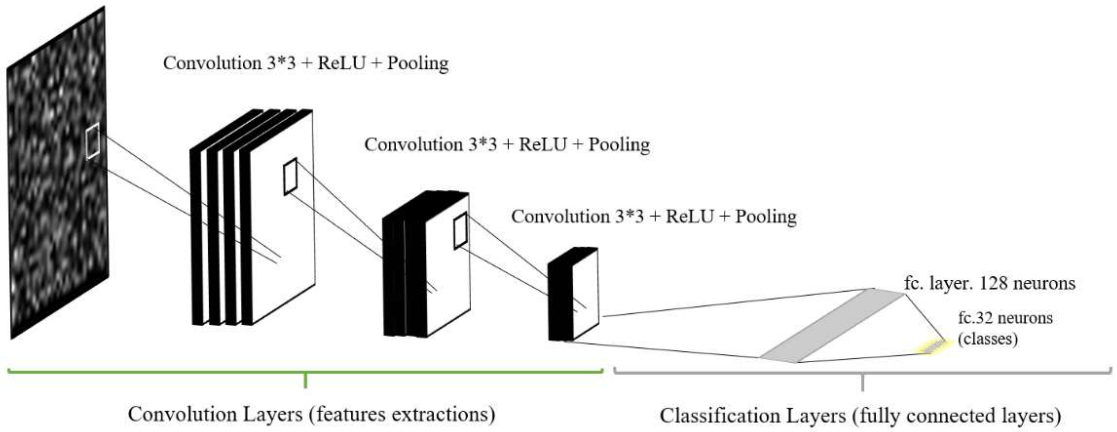

Fig. 4. Structure Of The Suggested Model For

recall are all important [43] are included depending on the confusion matrix computation. From Equation (2) to equation (10), the performance measures are shown.

$$\text{Testing Accuracy} = \frac{\text{TruePos+TrueNeg}}{(\text{TruePos+Fal} \quad)+(\text{TrueNeg+}False \quad)} \quad (7)$$

$$P = \frac{\text{TruePos}}{(\text{TruePos+False} \quad)} \quad (8)$$

$$R = \frac{\text{TruePos}}{(\text{TruePos+}False \quad)} \quad (9)$$

$$\text{F1 Score} = 2 * \frac{R*P}{(R+P)} \quad (10)$$

Where True-Pos indicates the overall value of True positive Samples collected, True-Neg shows the overall value of True-Neg samples collected, False-Pos indicates the overall number of False-Pos samples collected overall value, and False-Neg indicates the overall number of False-Neg samples.

4. RESULTS AND DISCUSSION

The experiments are carried out on a computer server with 96 GB of RAM and an Intel Xeon CPU, and MATLAB software (2 GHz) is used. During the experiments, the following specifications are chosen:

- For algorithms based on machine learning
  - Three different classifications are evaluated (decision trees, Ensemble and Support Vector Machine).
  -
  -
  - Two issues (treatment classification and treatment concentration prognosis).
  - The dataset is numerical form.
  - A 5-fold cross-validation is chosen.
  - As performance metrics, assessment accuracy, Receiver
  - Operating Characteristic (ROC), and Area Under Curve (AUC) are used.
- For DCNN
  - Utilizing the DCNN as suggested in Section 3.2.
  - Two issues (treatment classification and treatment concentration prognosis).
  - The dataset is in the form of a digital image.
  - The dataset was segmented into two halves (70 percent of the data for the training process, and 30 percent for the testing process).
  - To overcame categorization challenges in treatment data augmentation is employed.
  - Recall, F1 score, testing accuracy, and precision are utilized as performance indicators.

4.1 Treatment Classification Results

As per the subset picked from the original dataset, there are 32 treatment classes, which are listed in Table 2. Machine learning will be used to classify the treatments in a numerical format, and DCNN will be used in digital image format.

Table. 2 Classification Of Treatments Based On The Dataset

Thymoquinone	Ritonavir	Tenofovir Disoproxil Fumarate	solithromycin
Quinine hydrochloride	quinine-ethyl-carbonate	Remdesivir (GS-5734)	Ribavirin
Quinine	Pacritinib	Penciclovir	Polydatin
Indinavir	Nicotianamine	methylprednisolone-sodium-succinate	Imiquimod
Dimethyl fumarate	Hydroxychloroquine Sulfate	oseltamivir-carboxylate	GS-441524
CAL-101	Lopinavir	Haloperidol	Indomethacin
1-deoxygalactonojirimycin	Cobicistat	Darunavir	Chloroquine
Favipiravir	Aloxistatin	Camostat	Arbidol

The initial set of findings will be documented employing classical machine learning; three types of classical machine learning will be used: DT, SVM, and Ensemble. The mean testing accuracy applying for the chosen machine learning algorithm was shown in Table 3 utilizing 5k cross-validation.

Table 3. Accuracy Evaluation Utilizing A Machine Learning Technique With Family Algorithms

Family Algorithm	DT	SVM	Ensemble
Child-algorithm (finest-achieved accuracy)	Fine-Tree [44]	Cubic-SVM [45]	Subspace Discriminant [46]
Median Testing Accuracy	57.7%	71.5%	72.7%

The receiver operating characteristic curve (ROC curve) is among the efficient indicators for machine learning algorithms. A ROC diagram is a curve that depicts a classification model's performance across all categorization thresholds utilizing False Positive and True Positive Rates.

The ROC curves in Figure 5 illustrate the performance of various machine learning techniques on a single treatment of oseltamivir-carboxylate. The AUC is a composite metric that indicates effectiveness across all potential categorization criteria.

The AUC for therapy with oseltamivir-carboxylate was 73 percent when using DT, 84 percent when using SVM, and 86 percent when using Ensemble. Although experimental trials can provide approximately 96 ROC curves, the figures do not need to be repeated for various medications and Testing accuracy can be a useful measure of the machine learning algorithm quality.

The results obtained via the deep learning architecture usage are superior to those obtained by utilizing machine learning methods to enhance the efficiency and accuracy metrics of testing. By utilizing the proposed DCNN model and converting it to the picture domain via augmentation, the model was able to get superior outcomes. The testing accuracy achieved was 98.05 percent. Recall accuracy was 95.03 percent. The Precision was determined to be 96.52 percent accurate. The F1 score has an accuracy of 95.97 percent.

Figure 6 illustrates the confusion matrix. It is demonstrated unequivocally that utilizing a deep learning model with feature conversion to the image

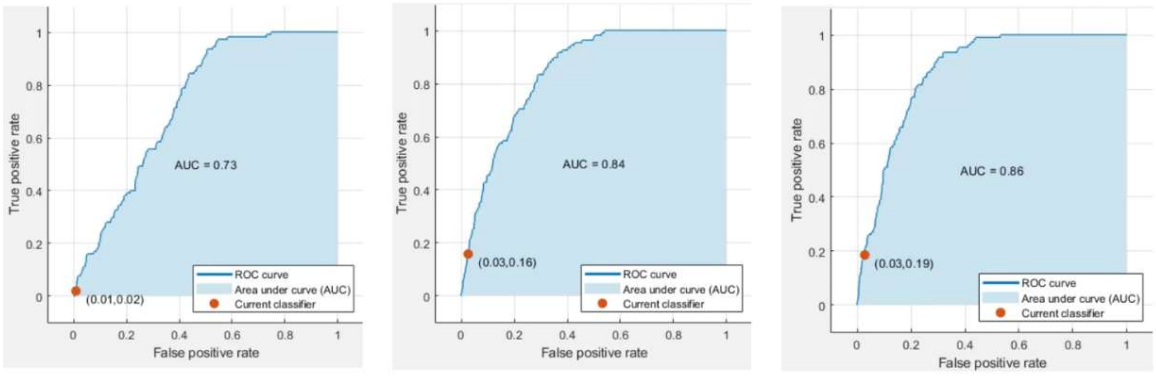

Fig 5. ROC Curves For The Combination Therapy Of Oseltamivir And Carboxylate Using (A) Decision Trees, (B) Support Vector Machines, And (C) Ensemble

Journal of Theoretical and Applied Information Technology
15th November 2021. Vol.99. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

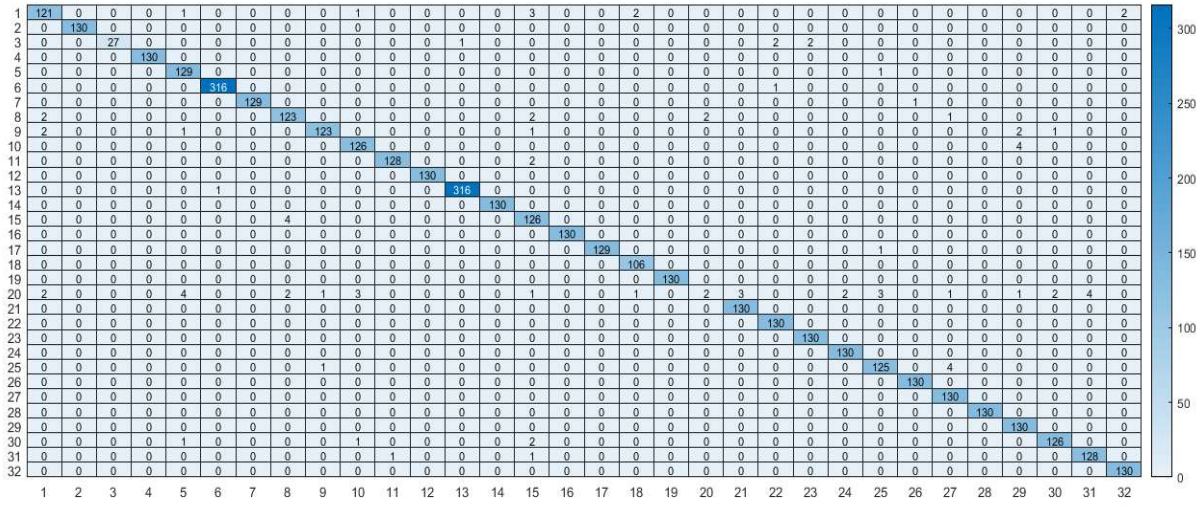

Fig 6. Matrix Of Confusion For The Suggested DCNN Model Dependent On Feature Images.

region increased testing accuracy by 25.35 percent when compared to an ensemble approach that scored 72.7 percent testing accuracy.

The evolution and progress of the suggested deep learning model's training phase is depicted in Figure 7, which depicts the training process progressing toward greater accuracy. The model has been tweaked to terminate training early if no improvement in accuracy is achieved after ten reiterations. 32 batches size were manufactured, and the rate of learning was 0.0001. Figure 8 illustrates instances of testing accuracy in conjunction with treatment classification.

Another objective of the suggested model is to forecast the treatment's concentration on the cell. The first step in determining the model's accuracy is to use a machine-learning algorithm to forecast the treatment concentration level. Three concentration levels were evaluated: 10%, 30%, and 100%. The accuracy of treatment concentration testing is shown in Table 4 utilizing DT, SVM, and ensemble algorithms with 5k cross-validation.

ROC curves and area under the curve (AUC) are further measures of a classifier's performance. Figure 9 shows the ROC curves for several machine learning techniques for the various categories of treatment concentrations of 10,30,100.

4.2 Results of Predicting Treatment Concentration

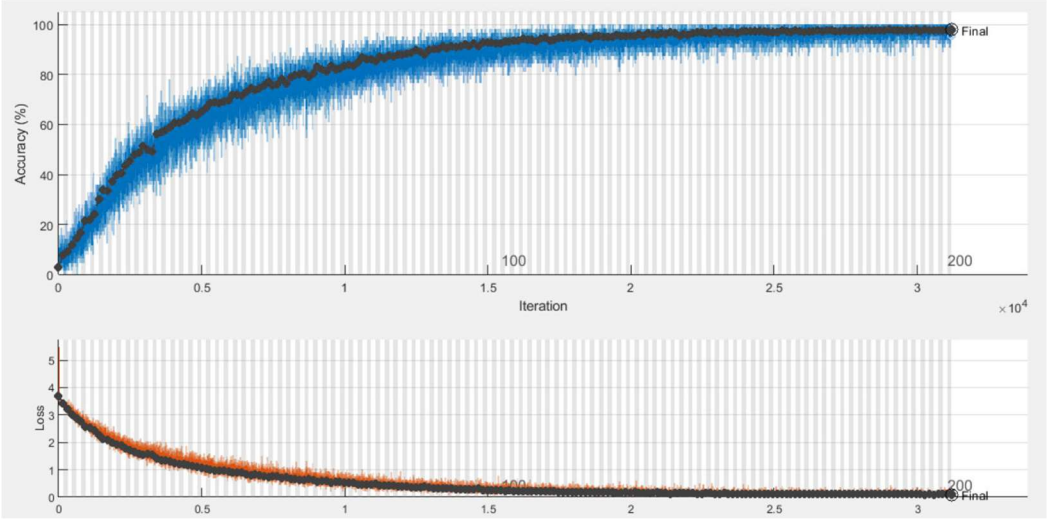

Fig 7. Accuracy Of Training With Loss Of Validation For The Proposed DCNN Model.

Journal of Theoretical and Applied Information Technology
15th November 2021. Vol.99. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

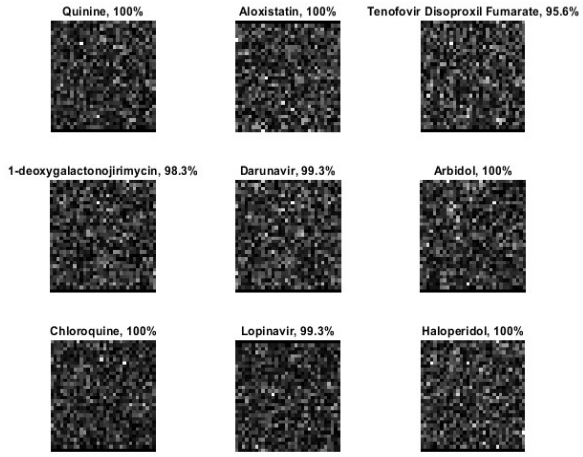

Fig 8. Illustrations Of Testing Accuracy For Categorization Of Treatments

The SVM and ensemble methods both obtained an AUC of 100%, which is an excellent sign of the classifier's quality. Additionally, as shown in Table 4, both classifiers (SVM and Ensemble) obtained 97.3 percent and 98.5 percent testing accuracy for a three-class problem, respectively.

Table 4. Accuracy Evaluation Utilizing A Variety Of Machine Learning Techniques

Family Algorithm	Ensemble	DT	SVM
Child-algorithm (highest level of accuracy)	Bagged tree [47]	Coarse Tree [48]	Linear-SVM [49]
Accuracy of Testing on the Average	98.5%	96.4%	97.3%

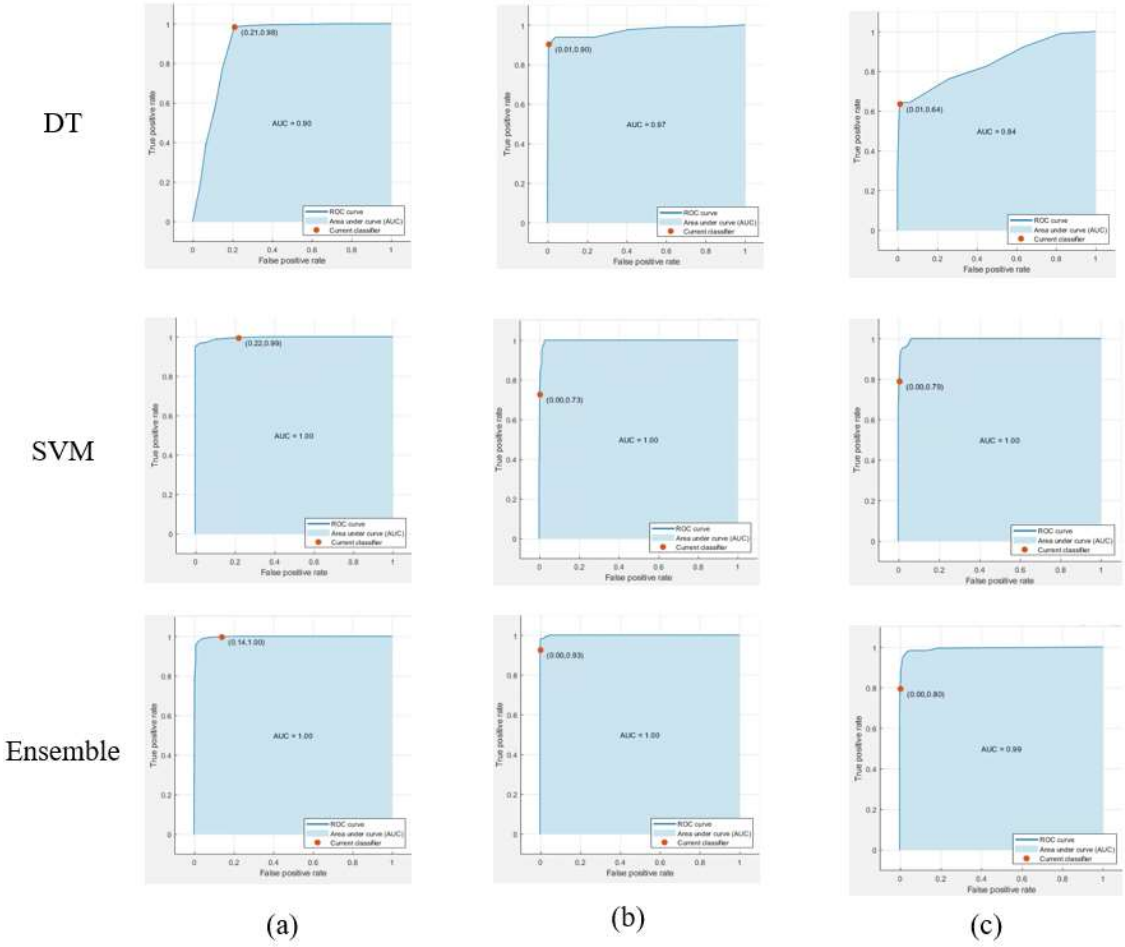

Fig 9. To Estimate The Medication Concentrations, AUC And ROC Of Machine Learning Algorithms Were Used For (A) 10, (B) 30, And (C) 100 Treatment Concentration Levels, Respectively.

The second approach is to address these issues utilizing deep learning, utilizing the same suggested DCNN model for digital image features without augmentation. There was no demand to utilize the augmentation procedure because the current proposal scored 98.2 percent testing accuracy. The matrix of confusion for the concentration ratios of the prospective medications is shown in Figure 10.

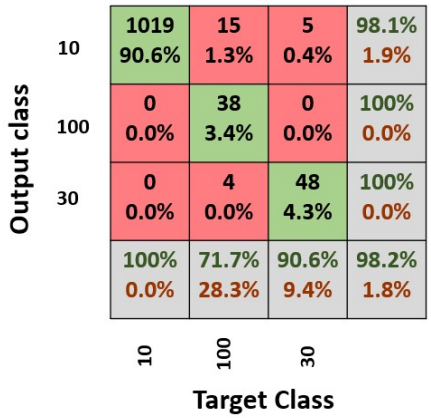

Fig 10. The Matrix Of Confusion For Predicting The Concentration Ratios Of Medication.

The suggested model with feature conversion obtained 98.2 percent testing accuracy and the following performance metrics: (Recall: 87.42 percent, Precision: 99.36 percent, and F1 score: 93.01 percent).

The scored accuracy was 98.1 percent for 10% of the concentration level and 100% for 30% of the concentration level. With a concentration ratio of 100%, the accuracy scored was also 100%. The scored accuracy for each category indicates the suggested DCNN model's performance.

5. DISCUSSION

When compared to machine learning techniques for treatment categorization, the suggested DCNN achieved a superior result in terms of testing accuracy when 32 classes are used. DCNN achieved 98.05 percent accuracy, while standard machine learning techniques, including such SVM, DT, and ensemble, obtained 57.7 percent, 71.5 percent, and 72.7 percent accuracy, consecutively. Performance measures corroborated the acquired findings for the suggested DCNN associated with characteristic conversion of images.

Conventional machine learning methods such as DT and SVM provided a near match to the suggested DCNN for predicting concentration levels of treatment. The SVM and DT obtained 97.3 and 96.4 percent accuracy in testing, consecutively, while the DCNN scored 98.2 percent accuracy. In order to assess accuracy, the ensemble approach surpassed the DCNN, attaining 98.5 percent. As a generalization, consider the standard machine learning approach for doing straightforward categorization problems, including such treatment concentration level prognosis, consisting of three categories. While performing multiclass classification tasks, including such treatment categorization, the deep learning model demonstrated superior efficiency and performance when compared to classical machine learning.

6. CONCLUSIONS AND FUTURE WORKS

COVID-19 epidemic has placed the world's healthcare systems in peril. There has been no therapy for this virus to date. One strategy for defeating this virus is to test licensed therapies on human cells as the first line of defense to shorten the time between medications and the revelation of a potential therapeutic. Deep learning and algorithms on computers can assist in closing that problem and aid in the search for a cure. The article objective is to describe a machine learning techniques and deep learning model for categorizing potential coronavirus therapies on a single human cell. The dataset utilized in this study is a variant of the online accessible RxRx.ai data. This study aims to develop an automated classification system for human cells based on their medications and medication concentration ratios. The DCNN model and approach provided here are dependent on arithmetic attributes conversion from the source dataset to the image domain. Two fully connected layers, three ReLU layers, three convolutional layers, three ReLU layers and three pooling layers comprise the suggested model. The research findings demonstrated that when compared to standard machine learning techniques, including such Support Vector Machine, Decision Tree, and Ensemble, the suggested DCNN model for medication categorization (32 class) scored 98.05 percent testing accuracy. The standard machine learning (Ensemble) algorithm scored 98.2 percent testing accuracy in treatment concentration level prognosis, whereas the suggested DCNN model reached 98.5 percent. One possible upcoming study is to replicate the studies using deep transfer models including such even deeper neural networks or Resnet50 and Alexnet to examine their efficiency on the dataset used in this study.

REFRENCES

[1] L. Chang, Y. Yan, and L. Wang, "Coronavirus Disease 2019: Coronaviruses and Blood Safety," *Transfusion Medicine Reviews*, 2020, doi: https://doi.org/10.1016/j.tmrv.2020.02.003.

[2] V. Chamola, V. Hassija, V. Gupta, and M. Guizani, "A Comprehensive Review of the COVID-19 Pandemic and the Role of IoT, Drones, AI, Blockchain, and 5G in Managing its Impact," *IEEE Access*, vol. 8, pp. 90225–90265, 2020, doi: 10.1109/ACCESS.2020.2992341.

[3] T. Singhal, "A Review of Coronavirus Disease-2019 (COVID-19)," *The Indian Journal of Pediatrics*, vol. 87, no. 4, pp. 281–286, 2020, doi: 10.1007/s12098-020-03263-6.

[4] M. Loey, F. Smarandache, and N. E. M. Khalifa, "Within the Lack of Chest COVID-19 X-ray Dataset: A Novel Detection Model Based on GAN and Deep Transfer Learning," *Symmetry*, vol. 12, no. 4, Art. no. 4, Apr. 2020, doi: 10.3390/sym12040651.

[5] C.-C. Lai, T.-P. Shih, W.-C. Ko, H.-J. Tang, and P.-R. Hsueh, "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges," *International Journal of Antimicrobial Agents*, vol. 55, no. 3, p. 105924, 2020, doi: https://doi.org/10.1016/j.ijantimicag.2020.105924.

[6] J. Li *et al.*, "Game consumption and the 2019 novel coronavirus," *The Lancet Infectious Diseases*, vol. 20, no. 3, pp. 275–276, Mar. 2020, doi: 10.1016/S1473-3099(20)30063-3.

[7] J. M. Sharfstein, S. J. Becker, and M. M. Mello, "Diagnostic Testing for the Novel Coronavirus," *JAMA*, Mar. 2020, doi: 10.1001/jama.2020.3864.

[8] "Countries where Coronavirus has spread - Worldometer." https://www.worldometers.info/coronavirus/countries-where-coronavirus-has-spread/ (accessed Jul. 10, 2020).

[9] M. Giovanetti, D. Benvenuto, S. Angeletti, and M. Ciccozzi, "The first two cases of 2019-nCoV in Italy: Where they come from?," *Journal of Medical Virology*, vol. 92, no. 5, pp. 518–521, May 2020, doi: 10.1002/jmv.25699.

[10] M. L. Holshue *et al.*, "First Case of 2019 Novel Coronavirus in the United States," *New England Journal of Medicine*, vol. 382, no. 10, pp. 929–936, Jan. 2020, doi: 10.1056/NEJMoa2001191.

[11] A. Khattar, P. R. Jain, and S. M. K. Quadri, "Effects of the Disastrous Pandemic COVID 19 on Learning Styles, Activities and Mental Health of Young Indian Students - A Machine Learning Approach," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, May 2020, pp. 1190–1195. doi: 10.1109/ICICCS48265.2020.9120955.

[12] C. Rothe *et al.*, "Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany," *New England Journal of Medicine*, vol. 382, no. 10, pp. 970–971, Jan. 2020, doi: 10.1056/NEJMc2001468.

[13] "Coronavirus (COVID-19) map." https://www.google.com/covid19-map/ (accessed Apr. 26, 2020).

[14] K. Ahmed, S. Abdelghafar, A. Salama, N. E. M. Khalifa, A. Darwish, and A. E. Hassanien, "Tracking of COVID-19 Geographical Infections on Real-Time Tweets." 2021.

[15] A. El-Aziz, A. Atrab, N. E. M. Khalifa, A. Darwsih, and A. E. Hassanien, "The role of emerging technologies for combating COVID-19 pandemic," in *Digital Transformation and Emerging Technologies for Fighting COVID-19 Pandemic: Innovative Approaches*, Springer, 2021, pp. 21–41.

[16] A. A. Abd El-Aziz, N. E. M. Khalifa, and A. E. Hassanien, "Exploring the Impacts of COVID-19 on Oil and Electricity Industry," in *The Global Environmental Effects During and Beyond COVID-19*, Springer, 2021, pp. 149–161.

[17] J. Civit-Masot, F. Luna-Perejón, M. Domínguez Morales, and A. Civit, "Deep Learning System for COVID-19 Diagnosis Aid Using X-ray Pulmonary Images," *Applied Sciences*, vol. 10, no. 13, Art. no. 13, Jan. 2020, doi: 10.3390/app10134640.

[18] A. Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman, and P. R. Pinheiro, "CovidGAN: Data Augmentation Using Auxiliary Classifier GAN for Improved Covid-19 Detection," *IEEE Access*, vol. 8, pp. 91916–91923, 2020, doi: 10.1109/ACCESS.2020.2994762.

[19] N. Narayan Das, N. Kumar, M. Kaur, V. Kumar, and D. Singh, "Automated Deep Transfer Learning-Based Approach for Detection of COVID-19 Infection in Chest X-

rays," *IRBM*, Jul. 2020, doi: 10.1016/j.irbm.2020.07.001.

[20] A. A. Ardakani, A. R. Kanafi, U. R. Acharya, N. Khadem, and A. Mohammadi, "Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks," *Computers in Biology and Medicine*, vol. 121, p. 103795, Jun. 2020, doi: 10.1016/j.compbiomed.2020.103795.

[21] M. Loey, A. ElSawy, and M. Afify, "Deep Learning in Plant Diseases Detection for Agricultural Crops: A Survey," *International Journal of Service Science, Management, Engineering, and Technology (IJSSMET)*, Apr. 01, 2020. www.igi-global.com/article/deep-learning-in-plant-diseases-detection-for-agricultural-crops/248499 (accessed Apr. 11, 2020).

[22] M. Loey, M. R. Naman, and H. H. Zayed, "A Survey on Blood Image Diseases Detection Using Deep Learning," *International Journal of Service Science, Management, Engineering, and Technology (IJSSMET)*, Jul. 01, 2020. www.igi-global.com/article/a-survey-on-blood-image-diseases-detection-using-deep-learning/256653 (accessed Jun. 17, 2020).

[23] N. Khalifa, M. Loey, M. Taha, and H. Mohamed, "Deep Transfer Learning Models for Medical Diabetic Retinopathy Detection," *Acta Informatica Medica*, vol. 27, no. 5, p. 327, 2019, doi: 10.5455/aim.2019.27.327-332.

[24] N. E. M. Khalifa, G. Manogaran, M. H. N. Taha, and M. Loey, "A deep learning semantic segmentation architecture for COVID-19 lesions discovery in limited chest CT datasets," *Expert Systems*, p. e12742.

[25] K. Heiser *et al.*, "Identification of potential treatments for COVID-19 through artificial intelligence-enabled phenomic analysis of human cells infected with SARS-CoV-2," *bioRxiv*, 2020.

[26] A. Çayir, I. Yenidoğan, and H. Dağ, "Feature Extraction Based on Deep Learning for Some Traditional Machine Learning Methods," in *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, Sep. 2018, pp. 494–497. doi: 10.1109/UBMK.2018.8566383.

[27] M. Jogin, Mohana, M. S. Madhulika, G. D. Divya, R. K. Meghana, and S. Apoorva, "Feature Extraction using Convolution Neural Networks (CNN) and Deep Learning," in *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)*, May 2018, pp. 2319–2323. doi: 10.1109/RTEICT42901.2018.9012507.

[28] A. Navada, A. N. Ansari, S. Patil, and B. A. Sonkamble, "Overview of use of decision tree algorithms in machine learning," in *2011 IEEE Control and System Graduate Research Colloquium*, Jun. 2011, pp. 37–42. doi: 10.1109/ICSGRC.2011.5991826.

[29] P.-L. Tu and J.-Y. Chung, "A new decision-tree classification algorithm for machine learning," in *Proceedings Fourth International Conference on Tools with Artificial Intelligence TAI '92*, Nov. 1992, pp. 370–377. doi: 10.1109/TAI.1992.246431.

[30] R. Polikar, "Ensemble Learning," in *Ensemble Machine Learning: Methods and Applications*, C. Zhang and Y. Ma, Eds. Boston, MA: Springer US, 2012, pp. 1–34. doi: 10.1007/978-1-4419-9326-7_1.

[31] I. Naseem, R. Togneri, and M. Bennamoun, "Linear Regression for Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 2106–2112, Nov. 2010, doi: 10.1109/TPAMI.2010.128.

[32] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, and M. Klein, *Logistic regression*. Springer, 2002.

[33] E. Mangalova and E. Agafonov, "Wind power forecasting using the k-nearest neighbors algorithm," *International Journal of Forecasting*, vol. 30, no. 2, pp. 402–406, Apr. 2014, doi: 10.1016/j.ijforecast.2013.07.008.

[34] Y. Xiao, J. Wu, Z. Lin, and X. Zhao, "A deep learning-based multi-model ensemble method for cancer prediction," *Computer Methods and Programs in Biomedicine*, vol. 153, pp. 1–9, 2018, doi: https://doi.org/10.1016/j.cmpb.2017.09.005.

[35] N. E. M. Khalifa, M. H. N. Taha, and A. E. Hassanien, "Aquarium family fish species identification system using deep neural networks," in *International Conference on Advanced Intelligent Systems and Informatics*, 2018, pp. 347–356.

[36] N. E. M. Khalifa, M. H. N. Taha, A. E. Hassanien, and A. A. Hemedan, "Deep bacteria: robust deep learning data augmentation design for limited bacterial colony dataset," *International Journal of Reasoning-based Intelligent Systems*, vol. 11, no. 3, pp. 256–264, 2019.

[37] N. E. M. Khalifa, M. H. N. Taha, D. E. Ali, A. Slowik, and A. E. Hassanien, "Artificial Intelligence Technique for Gene Expression by Tumor RNA-Seq Data: A Novel Optimized Deep Learning Approach," *IEEE Access*, vol. 8, pp. 22874–22883, 2020.

[38] M. Loey, M. Naman, and H. Zayed, "Deep Transfer Learning in Diagnosing Leukemia in Blood Cells," *Computers*, vol. 9, no. 2, Art. no. 2, Jun. 2020, doi: 10.3390/computers9020029.

[39] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, 2019.

[40] A. El-Sawy, M. Loey, and H. EL-Bakry, "Arabic Handwritten Characters Recognition Using Convolutional Neural Network," *WSEAS Transactions on Computer Research*, vol. 5, 2017, Accessed: Apr. 01, 2020. [Online]. Available: http://www.wseas.org/multimedia/journals/computerresearch/2017/a045818-075.php

[41] A. El-Sawy, H. EL-Bakry, and M. Loey, "CNN for Handwritten Arabic Digits Recognition Based on LeNet-5 BT - Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016," Cham, 2017, pp. 566–575.

[42] N. E. Khalifa, M. Loey, and S. Mirjalili, "A comprehensive survey of recent trends in deep learning for digital images augmentation," *Artificial Intelligence Review*, pp. 1–27, 2021.

[43] C. Goutte and E. Gaussier, "A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation," 2010. doi: 10.1007/978-3-540-31865-1_25.

[44] C. Barker, S. Cipkar, T. Lavigne, C. Watson, and M. Azzouz, "Real-Time Nuisance Fault Detection in Photovoltaic Generation Systems Using a Fine Tree Classifier," *Sustainability*, vol. 13, no. 4, p. 2235, 2021.

[45] U. Jain, K. Nathani, N. Ruban, A. N. J. Raj, Z. Zhuang, and V. G. Mahesh, "Cubic SVM classifier based feature extraction and emotion detection from speech signals," in *2018 International Conference on Sensor Networks and Signal Processing (SNSP)*, 2018, pp. 386–391.

[46] R. Hang, Q. Liu, H. Song, and Y. Sun, "Matrix-based discriminant subspace ensemble for hyperspectral image spatial–spectral feature fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 2, pp. 783–794, 2015.

[47] R. E. Banfield, L. O. Hall, K. W. Bowyer, and W. P. Kegelmeyer, "A comparison of decision tree ensemble creation techniques," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 1, pp. 173–180, 2006.

[48] T. Damrongsakmethee and V.-E. Neagoe, "Principal component analysis and relieff cascaded with decision tree for credit scoring," in *Computer Science On-line Conference*, 2019, pp. 85–95.

[49] Y.-W. Chang and C.-J. Lin, "Feature ranking using linear SVM," in *Causation and prediction challenge*, 2008, pp. 53–64.