A Decision Support System for Selecting the Most Suitable Machine Learning in Healthcare using User Parameters and Requirements.

Yashodhan Ketkar[a], Sushopti Gawade[b]

[a]*Department of Information Technology Engineering, Pillai College of Engineering, New Panvel, 410206, India*
[b]*Department of Computer Engineering, Pillai College of Engineering, New Panvel, 410206, India*

ARTICLE INFO

ABSTRACT

The use of machine learning in various fields is still limited. The driving reason behind this is the lack of easy-to-use systems for non-technical people. The objective of this paper is to provide the general population access to a machine learning system. We propose an automated machine learning system for non-technical users. The proposed system automates the selection of the best model as per user requirements. The proposed methods employ both the performance of the model and user requirements in the form of preferences for the model selection. The combination of these two features is calculated as the Vscores of the models. In this study, we employed the proposed system on Parkinson's disease datasets. The Support Vector Machine (SVM) and Random Forest (RF) are used as the algorithms for training the two models in the system. The system selected SVM as the best suited model for three test cases and the RF Model as the best suited model for one case. The proposed system showed high base accuracy in training the machine learning models and in the selection of appropriate models. Performed tests showed that the time parameter significantly affected the Vscore of the model.

1. Introduction

In recent years, machine learning has become very popular. Earlier, machine learning was limited to the research field. But quite recently, it has been used in various fields [39]. This is partially due to the increasing availability of computers and to the growth in computing power.

Researchers are using machine learning in various fields. One such field is solar energy production. In this field, the prediction of solar radiation available per day can be very significant. [1] employed a few machine learning algorithms for the prediction of the amount of solar radiation received in a day. Ağbulut et al. used natural properties such as weather, time, etc. as features for the machine learning algorithms. Ağbulut et al. concluded that supervised learning models gave satisfactory results in predicting the amount of solar radiation. Ağbulut et al. concluded that the Artificial Neural Network (ANN) model showed great potential for such specialized tasks.

In the field of chemistry, there is a wide range of factors that need to be considered. [50] used machine learning in the field of chemistry. Zuranski et al. suggest that the models performed better than conventional statistical methods. Zuranski et al. concluded that more research is necessary before implementation.

Machine learning can be applied to network security for intrusion detection. [30] state that the supervised learning methods perform satisfactorily for intrusion detection. Maseer et al. also remark that the supervised learning methods are limited to conventional problems.

[11] notes a lack of information and meaningful research in the medical field. Deo suggests that more support, subject-specific scope, and accuracy are important factors for the

higher impact of machine learning in the medical field. Deo also suggests that the ideal system should be able to take multiple data types for training.

[16] used machine learning with a combination of Internet of Things (IoT) for smart cities. Ghazal et al. conclude that machine learning shows promising results for smart cities. Ghazal et al. remark that machine learning is able to handle the high volume of data generated by sensors. [26] suggests that machine learning can handle downstream tasks efficiently.

This suggests that even with the immense popularity and accessibility of machine learning, it is still underutilized by the general population. This can be attributed to a lack of knowledge and skills. To solve this problem, we are proposing a system with minimum user interaction. The proposed system can be operated by users without prior knowledge of machine learning. This system allows them to train their machine learning algorithms.

The proposed system can be deployed on a local network. The system can feed data either manually or automatically according to user needs and policies. The local deployment also limits external access and reduces the influence of external factors on data.

2. Literature review

Supervised machine learning algorithms are widely used in various fields. The study conducted on the application of machine learning in various industries suggests that machine learning is already being utilized in various industries [6]. Supervised machine learning shows excellent results for well-labeled datasets. Such as in the case of Civil Structural Health Monitoring (SHM) machine learning, where supervised machine learning algorithms showed satisfactory results [13].

✉ ketkaryapr19me@student.mes.ac.in (Y. Ketkar); sgawade@mes.ac.in (S. Gawade)

Supervised learning algorithms are very efficient for conventional problems. The machine learning systems are capable of detecting Distributed Denial of Service (DDOS) attacks very accurately. The DDOS attack is a conventional network attack with various known features. Hence, the supervised algorithms are very successful in the prevention of such attacks in real time [35].

The study was conducted on the automated selection of support vector machine (SVM) algorithms with the goal of selecting the most optimal SVM model for the given tasks [24]. It was noted that the hyperparameters and the used data have an extreme impact on the prediction time of the algorithm.

Recently, machine learning was used in the diagnosis of Coronavirus Disease 2019 (COVID-19) disease and studies related to its structure. The survey of various studies was taken with various types of dataset [7]. These studies suggested that the supervised learning algorithms showed promising results in the diagnosis of COVID-19.

Machine learning is widely used in the diagnosis process, such as in the diagnosis of Parkinson's disease. In this study, a machine learning system used for the detection of Parkinson's disease achieved up to 100% accuracy with a confidence level higher than 95%. The study used Random Foreset (RF) and SVM algorithms [28]. The inclusion of machine learning in diagnosis leads to better results and higher accuracy while working with the high volume of data [15]. From a survey of over 200 published studies, it was noted that the machine learning system with supervised algorithms improved clinical decisions in Parkinson's disease diagnosis [31]. The study was conducted on the use of various machine learning models and resulted in SVM outperforming other models [36].

Arrhythmia is an irregular heartbeat, which is a common symptom of cardiological disorders. The supervised learning methods have already been used for the detection of arrhythmias with high accuracy and sensitivity [18]. Good feature selection and selection guidelines result in better performing machine learning systems [2]. With the help of the SVM model [40], 91.2% accuracy is obtained in the detection of arrhythmia.

The machine learning system needs to be robust, extremely accurate, and easy to use. The system that fulfills these essential requirements and provides early detection of arrhythmia is critical for better treatment of patients [41].

The Long Short Term Memory (LSTM), SVM, and Multiplayer Perceptron (MLP) models were used on the ECG signals of COVID-19 patients. These models were used with robots to successfully monitor the patient's data and condition. The SVM model produced satisfactory results [38].

The study of the automated selection of machine learning algorithms and their hyperparameters showed the limitations of the biomedical industry [27]. One of these limitations is the security and privacy aspects of the machine learning solution. While machine learning systems have great potential in healthcare systems, more research is needed to handle the security and privacy aspects [37].

A machine learning system that complies with Health Insurance Portability and Accountability Act (HIPPA) regulations was able to achieve up to 83% accuracy in the detection of heart disease [8]. A machine learning system with a multiperceptron neural network achieved up to 78% accuracy for stroke prediction with a relatively small dataset. The system was capable of producing even better results with large training datasets [12].

Machine learning utilizes various approaches to the same problem. In a study, two approaches were used to solve a diagnostic problem. In a direct approach, data is fed to a machine learning model. In an indirect approach, data is equalized before the machine learning process. The indirect method showed better results compared to the direct method. While the experiment was successful, it was noted that machine learning is still unstable for the medical field [22].

The supervised learning methods are extremely effective in the medical field and have proved useful in reducing the number of errors caused by humans [19]. The survey about the use of machine learning in the medical field suggested that big companies are already using the machine learning system for various tasks such as diagnosis and research. This introduced the need for a machine learning system that could be available to non-technical people. These systems should be able to handle various types of data [17].

There is a sufficient amount of literature available about automated machine learning systems. The automated system performed better than standalone systems, as observed from various benchmarks [49]. The use of an automated system reduces the complexity of the problems compared to regular machine learning systems [25]. The ability to handle high throughput of the data is another benefit of the automated system [23]. The automated systems are used as graph learning systems, with various frameworks already available in the industry [48]. The preprocessing pipelines in the automation system are able to handle large amounts of data as the results of tests improve with each iteration [32].

The automated system will allow non-technical people to use machine learning. The optimal SVM kernal selection with automated optimal hyperparameter calculations was performed with minimal hunman interaction. The system was able to calculate up to 2 hyperparameters with no human interaction [3]. Two other studies also focused on the selection of the optimal hyperparameters. In one of these studies, neural networks were used, while in the other study, a genetic algorithm was applied [45, 14]. Use of genetic algorithms was able to reduce uncertainty in the predictions of the system [14].

The automatic selection process can be extremely beneficial in dynamic environments. The excavation of soil or tunneling is one such environment. The displacement induced by excavation was predicted by machine learning with properties of soli as features. The unsupervised machine learning algorithm Genetic Algorithm Multilayer Perceptron (GA-MLP) showed good potential, while automated machine

learning was found to be the most optimal algorithm in these dynamic conditions [47]. It is noted that machine learning systems are extremely successful in dynamic environments [47, 29]. An automated machine learning system, used to enhance the academic performance of students, produced up to 75.9% accuracy. The system used data prior to the start of the academic year to predict the performance of the students. The system is able to detect the probability of failure with 83% accuracy [46].

There are multiple ways to select the best-suited model. The important part of such a system is the selection system. The system uses various approaches for selection; ranking is one of those approaches. The meta ranking is an alternative to the baseline ranking system. This system uses a multi-criteria ranking system for the selection of ideal classifiers [5].

Automated machine learning systems are already used in the medical field. Automated machine learning is employed for clinical note analysis [33] and the detection of breast cancer [34] by gene analysis. An automated system of twenty machine learning algorithms with graphical and web based interfaces for protein sequences has been deployed in one of the studies [44]. Recently, during the COVID-19 pandemic, an automated machine learning system was used for the purpose of screening. 98.3% accuracy was achieved during the screening [20]. The triage system and the survival rate prediction system were built with an automated machine learning system during the COVID-19 pandemic [9].

The civil engineering field also employed machine learning systems. An automated system was used during the disaster planning by studying the complex dynamics of glabal landslides [42]. Another automated system, which was deployed to automate the workflow, performed consistently and objectively [43]. An automated machine learning system was used to observe the physical and hydraulic soils for an accurate measurement of the soil moisture content. This system was able to extract the features from the data without human interaction [4].

3. Design and implmentation

3.1. Mathematical model

The system is able to work with multiple models. So let's assume we have N number of models, Model 1, Model 2,···, Model N for the task. Each model has six performance parameters, they are Accuracy P_1, F1 Score P_2, Precision P_3, Recall P_4, Area under ROC (Receiver Operating Characteristic) P_5 and Prediction Time P_6. Each parameter is assigned a predefined weight depending on the user's requirements. These performance parameters and weights are used to evaluate the total performance score of models. The performance score of the model is obtained by subtracting the weighted time parameter from the summation of weighted parameters.

$$V_{score} = \left(\sum_{x=1}^{5} w_x P_x \right) - w_6 P_6 \qquad (1)$$

The Vscore of each model is calculated with eq. (1). For a given N number of models, we will get V_1, V_2, ···, V_n. The model with the highest Vscore will be selected as the best-suited model.

3.1.1. Weightage

The weights in the system are selected by the user's requirements. The weights are calculated by the system. The value of weight lies within the range of 0.2 and 1. The default weights for the first five parameters are set to 0.6. The weights of parameters are dependent on each other, except for the sixth (time) parameter, which is independent of other parameters. In the case of the sixth parameter, the default weight is set at 0.5. The value can be changed to either 0.25 or 0.75 with respect to the user's requirements.

3.2. System architecture

The goal of the system is to be used by non-technical users. Figure 1 shows the mathematical model and system architecture in detail. As shown in the figure, part of the user requirements are used for the generation of models. This information contains the model templates along with predermined hyperparameters for model generation. With this template, the models are generated and trained with 80% of the training dataset. The remaining 20% of the dataset is used for the testing process for the extraction of performance parameters. As mentioned earlier, six parameters are used in this system. These parameters are forwarded to the Vscore calculation process. In this process, the user requirements are used to generate the weightage for each parameter. These weighted parameters are used to calculate the Vscore mentioned earlier. In the next step, these Vscores are compared, and the model with the higher Vscore is selected as the most suited model. The selected model is stored in the system for future use along with the performance parameters of the model.

3.3. System processes

The system consists of three processes, or modules. These processes are the training process, selection process, and prediction process. The first two processes interact with each other. Their task is to produce the best-suited model for the user's needs. The third process interacts with the selected models to generate the predictions. The user is indirectly allowed to access the training and prediction process. The user data is stored on a local drive for easier and faster access. The following subsection will give a further explanation of these processes.

3.3.1. Training process

The training process is the first process in the system. Figure 2 shows the structure of the training process. The training process has many functions. The first function is gathering data from the user and preprocessing it for training purposes. The training process also generates models with the template, which contains model structure and parameters. These models are trained with processed data and stored
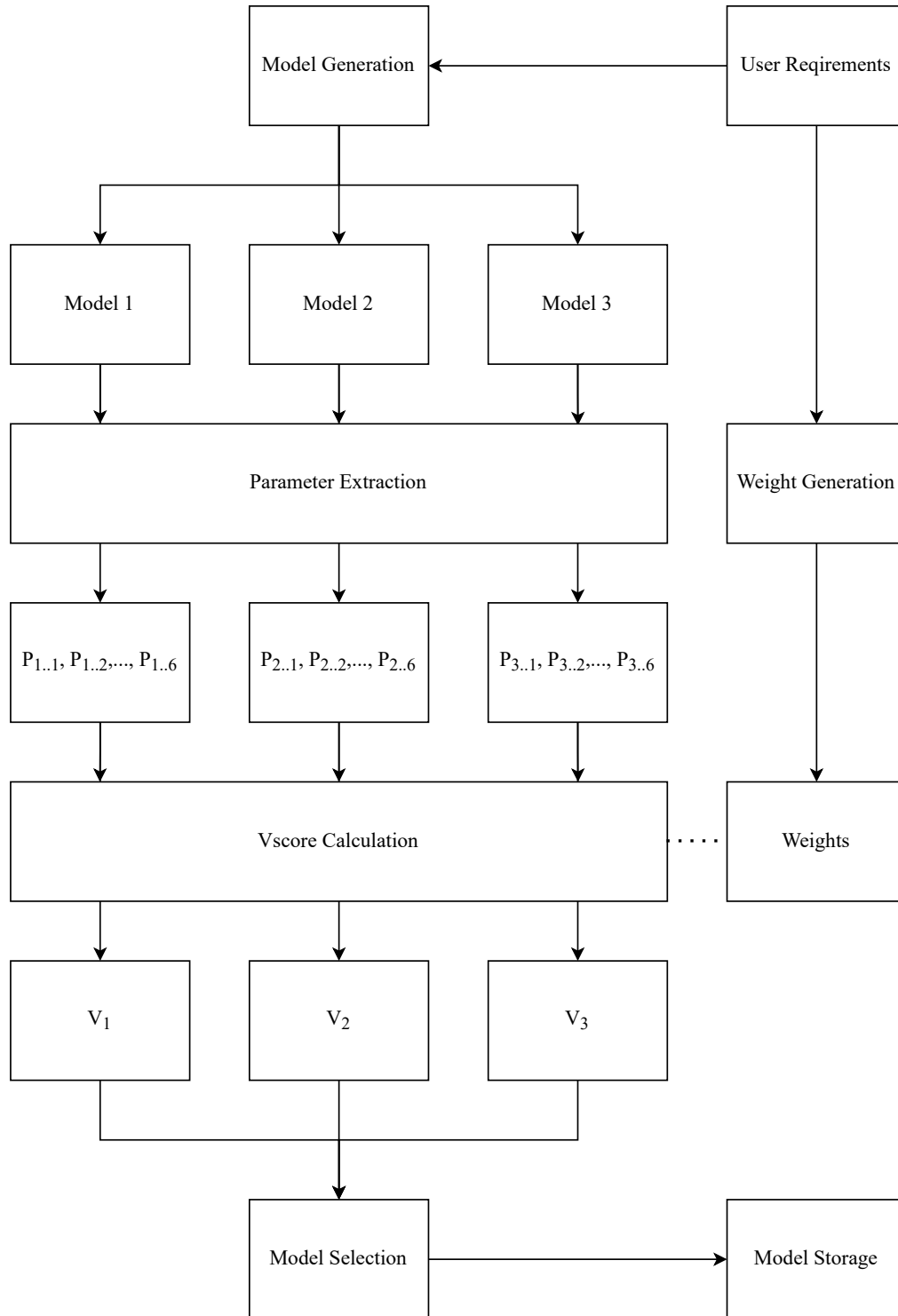
```
┌──────────────────┐                          ┌──────────────────┐
│ Model Generation │ ◄────────────────────────│ User Reqirements │
└──────────────────┘                          └──────────────────┘
```

Figure 1: Model Process

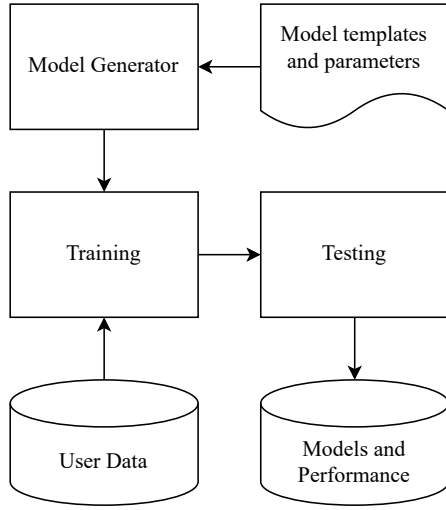for future use. The performance of models is also calculated during this phase and stored for selection purposes.

Figure 2: Training Process

3.3.2. *Selection process*

The selection process is second in the system. Figure 3 shows the structure of the training process. This process evaluates the performance of models based on metrics and weightage. The metrics of models are generated during the training process. The performance weightage is defined by the user depending on requirements. The final performance score is calculated and used for the selection of the best model for user's task. The selected model is stored in a separate directory with a label for easier access. This model will be used for prediction problems.
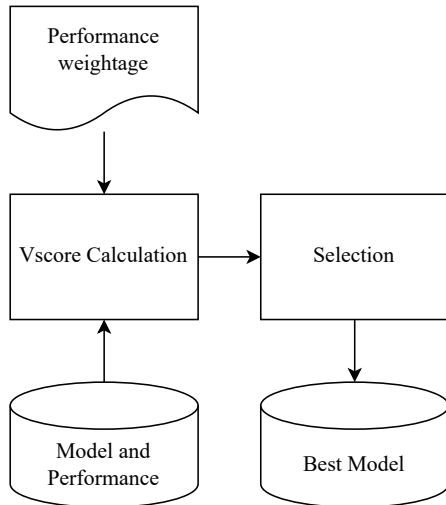

Figure 3: Selection Process

3.3.3. *Prediction process*

The prediction process is the final process in the system. Figure 4 shows the structure of the prediction process. This

process unpacks the best model and loads it for prediction. The model generates predictions with user-provided data. The output is displayed to the user. This output is also stored for future reference.
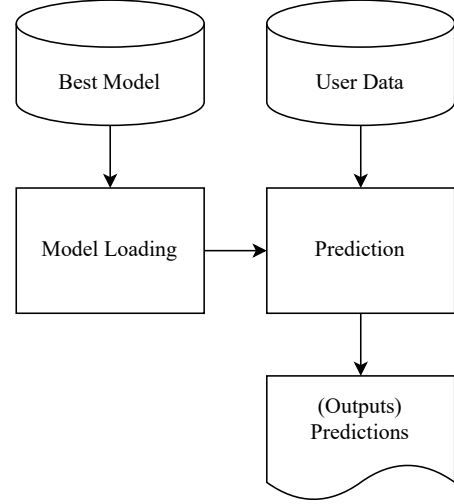

Figure 4: Prediction Process

3.4. Algorithms

The system uses two algorithms to run. These algorithms are training and selection algorithms and prediction algorithms.

3.4.1. *Training and selection algorithm*

1. Collect and receive dataset.

2. Split data into an 80:20 ratio for training and testing.

3. Build a model from presets.

4. Train models with training datasets and store models.

5. Evaluate the performance of models with the testing dataset.

6. Rank models with the help of performance and premade tuning parameters.

7. Select the best model and store it for future use.

3.4.2. *Prediction algorithm*

1. Collect and receive dataset.

2. Then load the best-suited model data from the storage.

3. Unpack the model for predictions.

4. Make predictions with the provided dataset and loaded model.

5. Return predictions to the user.

3.5. Implmentation

3.5.1. *Web architecture*

The system provides service with a web application. The users aren't allowed to interact with the system directly. This provision is to provide security and reduce the outside influence on results. The interface layer is used for a user to interact with the system indirectly. Figure 5 shows the web architecture of the system. The system is connected to the database directly. A direct connection is provided to access live data. The user is provided with a list of uploaded (user uploaded) datasets along with the preference options. The user selects a training dataset from this list and provides preference according to its needs. A user-selected dataset is processed by the system and stored in the database. The user provided preferences are used for the weightage generation. Currently, the system only allows staff members to access the results and patient data, and training options are restricted to people with admin roles.
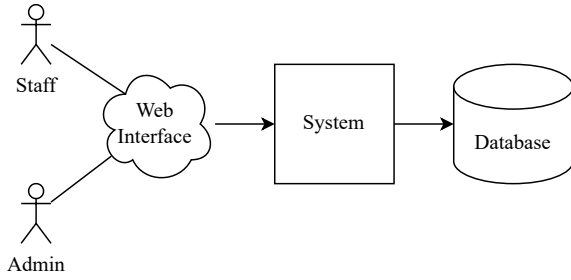

Figure 5: Web Architecture

3.5.2. *Web interface*

The web interface provides users with a way to interact with the system. The primary function of the web interface is to provide secure access to the system process. Approved users can interact with various modules on the system. This interface allows users to access selection and prediction processes. Users can upload data to the system.

The interface presents prediction results as well as performance evaluation results. Prediction results are provided in tabular format and stored in Comma Separated Values (CSV) files for future reference. Performance results are presented in graphical format and also stored in CSV files for future reference.

4. Result and analysis

4.1. Dataset and test cases

To test the selection system we are employing two models and one dataset. The dataset contains 201 records with 754 features [10]. The random forest algorithm and support vector machine algorithm is used to generate the models $Model_1$ and $Model_2$ respectively. The performance obtained from these two models is shown in table 1. The table shows the base performance of two models. This base performance will be weighted to calculate the Vscore of the models.

We will use 4 cases to test the system. The test cases are:

Table 1
Performance Analysis of Dataset

Performance Parameter		$Model_1$	$Model_2$
P_1	Accuracy	0.8039	0.7451
P_2	F1 Score	0.8750	0.8434
P_3	Pricision	0.8537	0.7955
P_4	Recall	0.8974	0.8974
P_5	Area under ROC	0.6987	0.5737
P_6	Prediction Time	0.5563	0.0981

Table 2
Weightage for cases

Weights	$Case_1$	$Case_2$	$Case_3$	$Case_4$
w_1	0.6	1.0	1.0	1.0
w_2	0.6	0.4	0.4	0.4
w_3	0.6	0.8	0.8	0.8
w_4	0.6	0.2	0.2	0.2
w_5	0.6	0.6	0.6	0.6
w_6	0.50	0.25	0.50	0.75

Case I: We did not provide any preference in the selection parameters. Hence the default weight for all parameters. The default weightage of the system is 0.6 for the first five parameters and 0.5 for the time parameter.

Case II: We gave preference to parameters in this order: 1. Accuracy (1.0) 2. Pricision (0.8) 3. Area under ROC (0.6) 4. F1 Score (0.4) 5. Recall (0.2).
Also, the sixth parameter (prediction time) is set to careful approach (0.25).

Case III: In this case, we used similar preferences from previous case except for the sixth parameter, where we selected default approach (0.50).

Case IV: Similar to the previous case we only changed the sixth parameter to the faster approach (0.75).

Table 2, shows the weight values for each case. These values are calculated with the help of the priority set by users. These values range between 0.2 and 1.0, and the sum of the first five weights is equal to 3.0.

4.2. Test results

We used eq. (1) to calculate the Vscore of the models. Table 3, shows the Vscore of $Model_1$ and $Model_2$ values along with the model selected by the system.

As seen in table 3, $Model_2$ is selected as the best model in three out of four cases. While $Model_1$ is selected as the best model is Case II, where we gave preference to a careful approach. The time parameter played an important role in the selection of the models compared to the other five parameters. Hence, implementation of the time parameter proved beneficial in the final selection of the model. Figure 6 shows this result (Vscores) in graphical format.
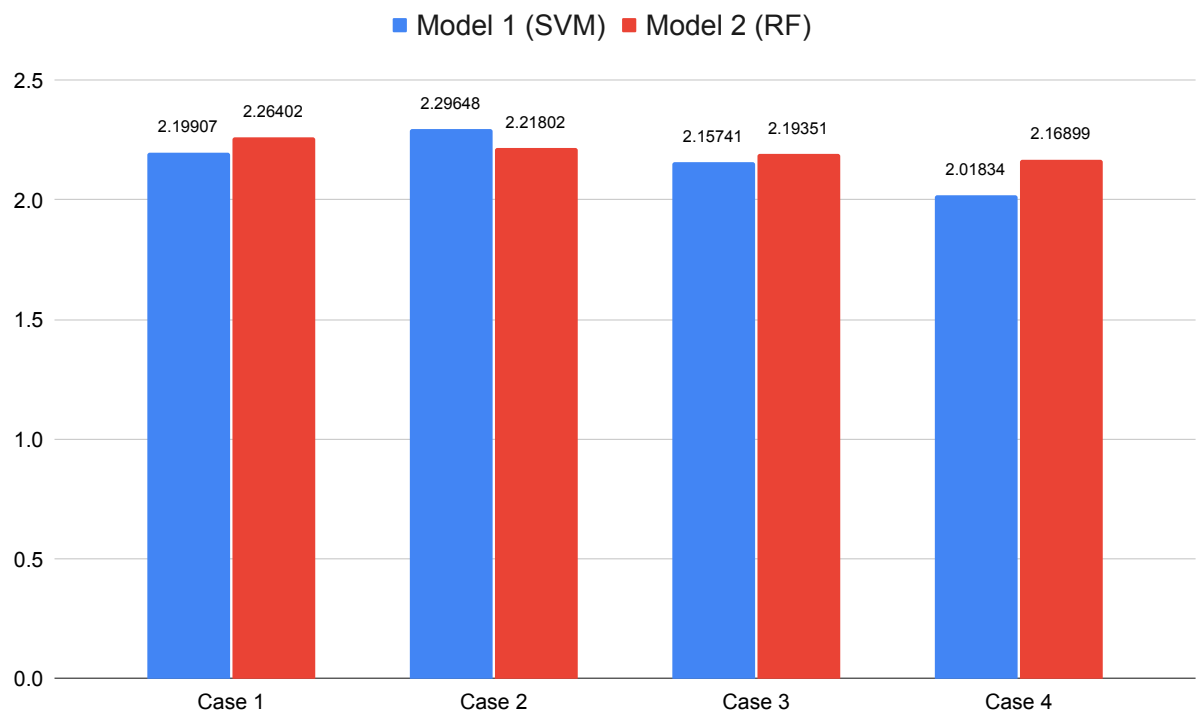
Figure 6: Vscore of test cases

Table 3
Results of test cases

Case	V_{score}|Model$_1$	V_{score}|Model$_2$	Model Selected
I	2.19907	2.26402	Model$_2$
II	2.29648	2.21802	Model$_1$
III	2.15741	2.19351	Model$_2$
IV	2.01834	2.16899	Model$_2$

4.2.1. *Key findings*

From the data displayed in previous section, we saw how the system works on different datasets. These are a few key findings we obtained from that knowledge.

1. The system successfully works with two different models.
2. The scale of data does impact the efficiency and effectiveness of the system.
3. The system successfully uses the weightage generated from the user requirements.
4. The weightage of the user, especially the weight for time parameter significantly impacts the decision of the system.

5. Novelty of the work

The suggested system in this paper is a weightage based model selection system. The novelty of the system is the combination of the user selected preferences and the performance parameters of the models for the selection of the most suited model. The system is able to handle a large number of models, which are generated by the templates provided by the user. The suggested system can work automatically with minimal human interaction. The user is only required to provide the training data to the system. The system will use this data to select the most suitable model. The user can use options to select preferred parameters, otherwise default values will be used in the selection process.

6. Conclusion and future work

The system performed satisfactorily during the tests. The application was able to select the model for the provided dataset. The best-suited model was able to meet the user requirements. The whole process required minimum human interaction. The system can be used with any number of models. It allows users with limited prior knowledge easier access to machine learning technology. The user-defined parameter weights lead to the selection of the best-suited model for particular tasks. The performance of models is stored for future evaluation of the system.

Currently, the system is limited to only supervised machine learning algorithms. The supervised nature of these algorithms limits the training dataset to the labeled dataset. By providing support to unsupervised learning algorithms, systems can accommodate various types of training datasets. The selection parameters are adjusted before the training process. These predefined parameters restrict the selection

choices of the system. Allowing users to tweak selection parameters after the training process will allow them to meet user requirements more efficiently. Future work will also focus on the implementation of Robotic Process Automation (RPA) tools for the data collection for easier integration with the old system [21]. Future work will also focus on comparing the performance between the proposed system and similar automation systems.

References

[1] Ağbulut, Ü., Gürel, A.E., Biçen, Y., 2021. Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison. Renewable and Sustainable Energy Reviews 135, 110114.

[2] Alfaras, M., Soriano, M.C., Ortín, S., 2019. A fast machine learning model for ecg-based heartbeat classification and arrhythmia detection. Frontiers in Physics 7, 103.

[3] Ayat, N.E., Cheriet, M., Suen, C.Y., 2005. Automatic model selection for the optimization of svm kernels. Pattern Recognition 38, 1733–1745.

[4] Babaeian, E., Paheding, S., Siddique, N., Devabhaktuni, V.K., Tuller, M., 2021. Estimation of root zone soil moisture from ground and remotely sensed soil information with multisensor data fusion and automated machine learning. Remote Sensing of Environment 260, 112434.

[5] Brazdil, P.B., Soares, C., Da Costa, J.P., 2003. Ranking learning algorithms: Using ibl and meta-learning on accuracy and time results. Machine Learning 50, 251–277.

[6] Burkart, N., Huber, M.F., 2021. A survey on the explainability of supervised machine learning. Journal of Artificial Intelligence Research 70, 245–317.

[7] Chadaga, K., Prabhu, S., Vivekananda, B.K., Niranjana, S., Umakanth, S., 2021. Battling covid-19 using machine learning: A review. Cogent Engineering 8, 1958666.

[8] Chang, V., Bhavani, V.R., Xu, A.Q., Hossain, M., 2022. An artificial intelligence model for heart disease detection using machine learning algorithms. Healthcare Analytics 2, 100016.

[9] Chen, Z., Zhao, P., Li, C., Li, F., Xiang, D., Chen, Y.Z., Akutsu, T., Daly, R.J., Webb, G.I., Zhao, Q., et al., 2021. ilearnplus: a comprehensive and automated machine-learning platform for nucleic acid and protein sequence analysis, prediction and visualization. Nucleic acids research 49, e60–e60.

[10] Debasis, S., (2020, Oct). Parkinson Disease Detection, Version 2. https://www.kaggle.com/datasets/debasisdotcom/parkinson-disease-detection?sort=votes. Accessed on 20 Apr 2022.

[11] Deo, R.C., 2015. Machine learning in medicine. Circulation 132, 1920–1930.

[12] Dev, S., Wang, H., Nwosu, C.S., Jain, N., Veeravalli, B., John, D., 2022. A predictive analytics approach for stroke prediction using machine learning and neural networks. Healthcare Analytics 2, 100032.

[13] Flah, M., Nunez, I., Ben Chaabene, W., Nehdi, M.L., 2021. Machine learning algorithms in civil structural health monitoring: a systematic review. Archives of computational methods in engineering 28, 2621–2643.

[14] Garcia, L.G., Lôndero, V., 2021. A parameter optimizer based on genetic algorithm for the simulation of carbonate facies. Intelligent Systems with Applications 12, 200057.

[15] Ghaderzadeh, M., Asadi, F., Hosseini, A., Bashash, D., Abolghasemi, H., Roshanpour, A., 2021. Machine learning in detection and classification of leukemia using smear blood images: a systematic review. Scientific Programming 2021.

[16] Ghazal, T.M., Hasan, M.K., Alshurideh, M.T., Alzoubi, H.M., Ahmad, M., Akbar, S.S., Al Kurdi, B., Akour, I.A., 2021. Iot for smart cities: Machine learning approaches in smart healthcare—a review. Future Internet 13, 218.

[17] Greener, J.G., Kandathil, S.M., Moffat, L., Jones, D.T., 2022. A guide to machine learning for biologists. Nature Reviews Molecular Cell Biology 23, 40–55.

[18] Guvenir, H.A., Acar, B., Demiroz, G., Cekin, A., 1997. A supervised machine learning algorithm for arrhythmia analysis, in: Computers in Cardiology 1997, IEEE. pp. 433–436.

[19] Ibrahim, I., Abdulazeez, A., 2021. The role of machine learning algorithms for diagnosing diseases. Journal of Applied Science and Technology Trends 2, 10–19.

[20] Ikemura, K., Bellin, E., Yagi, Y., Billett, H., Saada, M., Simone, K., Stahl, L., Szymanski, J., Goldstein, D., Gil, M.R., et al., 2021. Using automated machine learning to predict the mortality of patients with covid-19: prediction model development study. Journal of medical Internet research 23, e23458.

[21] Ketkar, Y., Gawade, S., 2021. Effectiveness of robotic process automation for data mining using uipath, in: 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), pp. 864–867. doi:10.1109/ICAIS50930.2021.9396024.

[22] Kim, I.K., Lee, K., Park, J.H., Baek, J., Lee, W.K., 2021. Classification of pachychoroid disease on ultrawide-field indocyanine green angiography using auto-machine learning platform. British Journal of Ophthalmology 105, 856–861.

[23] Koh, J.C., Spangenberg, G., Kant, S., 2021. Automated machine learning for high-throughput image-based plant phenotyping. Remote Sensing 13, 858.

[24] Lee, J.H., Lin, C.J., 2000. Automatic model selection for support vector machines. CSIE, NTU .

[25] Li, K.Y., Burnside, N.G., de Lima, R.S., Peciña, M.V., Sepp, K., Cabral Pinheiro, V.H., de Lima, B.R.C.A., Yang, M.D., Vain, A., Sepp, K., 2021. An automated machine learning framework in unmanned aircraft systems: new insights into agricultural management practices recognition approaches. Remote Sensing 13, 3190.

[26] Liu, X., Zhang, F., Hou, Z., Mian, L., Wang, Z., Zhang, J., Tang, J., 2021. Self-supervised learning: Generative or contrastive. IEEE Transactions on Knowledge and Data Engineering .

[27] Luo, G., 2016. A review of automatic selection methods for machine learning algorithms and hyper-parameter values. Network Modeling Analysis in Health Informatics and Bioinformatics 5, 1–16.

[28] Mandal, I., Sairam, N., 2014. New machine-learning algorithms for prediction of parkinson's disease. International Journal of Systems Science 45, 647–666.

[29] Maschler, B., Weyrich, M., 2021. Deep transfer learning for industrial automation: a review and discussion of new techniques for data-driven machine learning. IEEE Industrial Electronics Magazine 15, 65–75.

[30] Maseer, Z.K., Yusof, R., Bahaman, N., Mostafa, S.A., Foozy, C.F.M., 2021. Benchmarking of machine learning for anomaly based intrusion detection systems in the cicids2017 dataset. IEEE access 9, 22351–22370.

[31] Mei, J., Desrosiers, C., Frasnelli, J., 2021. Machine learning for the diagnosis of parkinson's disease: A review of literature. Frontiers in aging neuroscience 13, 184.

[32] Mohr, F., Wever, M., Tornede, A., Hüllermeier, E., 2021. Predicting machine learning pipeline runtimes in the context of automated machine learning. IEEE Transactions on Pattern Analysis and Machine Intelligence 43, 3055–3066.

[33] Mustafa, A., Rahimi Azghadi, M., 2021. Automated machine learning for healthcare and clinical notes analysis. Computers 10, 24.

[34] Panagopoulou, M., Karaglani, M., Manolopoulos, V.G., Iliopoulos, I., Tsamardinos, I., Chatzaki, E., 2021. Deciphering the methylation landscape in breast cancer: Diagnostic and prognostic biosignatures through automated machine learning. Cancers 13, 1677.

[35] Pande, S., Khamparia, A., Gupta, D., Thanh, D.N., 2021. Ddos detection using machine learning technique, in: Recent Studies on Computational Intelligence. Springer, pp. 59–68.

[36] Prashanth, R., Roy, S.D., Mandal, P.K., Ghosh, S., 2016. High-accuracy detection of early parkinson's disease through

multimodal features and machine learning. International journal of medical informatics 90, 13–21.

[37] Qayyum, A., Qadir, J., Bilal, M., Al-Fuqaha, A., 2020. Secure and robust machine learning for healthcare: A survey. IEEE Reviews in Biomedical Engineering 14, 156–180.

[38] Rehmat, M.A., Hassan, M.A., Khalid, M.H., Dilawar, M., 2022. Next level of hospitalisation through smart icu. Intelligent Systems with Applications 14, 200080.

[39] Sarker, I.H., 2021. Machine learning: Algorithms, real-world applications and research directions. SN Computer Science 2, 1–21.

[40] Shimpi, P., Shah, S., Shroff, M., Godbole, A., 2017. A machine learning approach for the classification of cardiac arrhythmia, in: 2017 international conference on computing methodologies and communication (ICCMC), IEEE. pp. 603–607.

[41] Soman, T., Bobbie, P.O., 2005. Classification of arrhythmia using machine learning techniques. WSEAS Transactions on computers 4, 548–552.

[42] Sufi, F.K., Alsulami, M., 2021. Knowledge discovery of global landslides using automated machine learning algorithms. IEEE Access 9, 131400–131419.

[43] Sun, A.Y., Scanlon, B.R., Save, H., Rateb, A., 2021. Reconstruction of grace total water storage through automated machine learning. Water Resources Research 57, e2020WR028666.

[44] Tran, N.K., Howard, T., Walsh, R., Pepper, J., Loegering, J., Phinney, B., Salemi, M.R., Rashidi, H.H., 2021. Novel application of automated machine learning with maldi-tof-ms for rapid high-throughput screening of covid-19: A proof of concept. Scientific reports 11, 1–10.

[45] Yang, D., Myronenko, A., Wang, X., Xu, Z., Roth, H.R., Xu, D., 2021. T-automl: Automated machine learning for lesion segmentation using transformers in 3d medical imaging, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3962–3974.

[46] Zeineddine, H., Braendle, U., Farah, A., 2021. Enhancing prediction of student success: Automated machine learning approach. Computers & Electrical Engineering 89, 106903.

[47] Zhang, D., Shen, Y., Huang, Z., Xie, X., 2022. Auto machine learning-based modelling and prediction of excavation-induced tunnel displacement. Journal of Rock Mechanics and Geotechnical Engineering .

[48] Zhang, Z., Wang, X., Zhu, W., 2021. Automated machine learning on graphs: A survey. arXiv preprint arXiv:2103.00742 .

[49] Zöller, M.A., Huber, M.F., 2021. Benchmark and survey of automated machine learning frameworks. Journal of artificial intelligence research 70, 409–472.

[50] Zuranski, A.M., Martinez Alvarado, J.I., Shields, B.J., Doyle, A.G., 2021. Predicting reaction yields via supervised learning. Accounts of chemical research 54, 1856–1865.