Journal of Information Systems Engineering and Management
2025, 10(9s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Hybrid Bio-Inspired GRU Model with Multiheaded Attention for Predicting Smart Grid Stability

Kazi Kalim Nasir[1], Dr. Jitendra N Shinde[2], Dr. Raju M Sairise[3]
[1]*Research Scholar, Department of Electrical Engineering, Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu, Rajasthan, India. kazikalim00@gmail.com*
[2]*Associate Professor, Department of Electrical Engineering, Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu, Rajasthan, India. jn_shinde@yahoo.com*
[3]*Principal & Associate Professor, Yadavrao Tasgaonkar College of Engineering and Management, Raigad, Maharashtra, India. rsairise566@tasgaonkartech.com.*

ARTICLE INFO

Received: 21 Oct 2024

Revised: 15 Dec 2024

Accepted: 30 Dec 2024

ABSTRACT

Smart grids are a big part of modern energy systems because they make it easier to handle and distribute energy. Still, it's hard to keep the grid stable because working conditions, supply, and demand for energy are always changing. The Bioinspired (PSO) + GRU Model is a strong way to handle the complexity of grid stability prediction that is shown in this work. An algorithm based on biology and deep learning make the system very accurate and reliable. It uses Gated Recurrent Units (GRUs) to model time and Particle Swarm Optimisation (PSO) to choose which features to use. The suggested design uses the Smart Grid Stability collection, which has 60,000 records of grid properties such as delay, flexibility, and power. It also has a goal variable that shows how stable the grid is. Using tools like SMote can help even out the classes in a dataset, which makes sure that the model is trained fairly. PSO is used to pick the most important features, which lowers the number of dimensions and speeds up processing. GRUs can find sequential links in the grid's operational data using Multiheaded attention methods. This lets them make accurate predictions about stability. For example, SVC, LGBR, and ANN are not as accurate as this model, which has an F1-score of 99.1% and an accuracy rate of 99.5%. The confusion matrix study shows that the framework is even more stable, with low judgement mistakes. Thorough planning steps, such as normalisation, dimensionality reduction, data cleaning, and more, make sure that the model gets good inputs. This paper stresses how important it is for smart grid systems to use cutting edge deep learning and efficiency methods to solve real-world problems. The suggested method could be used to make predictions more accurate and could also be scaled up to be used in real-time grid tracking systems. In the future, this method could be used in other study areas that need model time data and big decisions.

Keywords: Smart Grid Stability, Bioinspired Optimization, Particle Swarm Optimization (PSO), Gated Recurrent Unit (GRU), Deep Learning, Temporal Data Modeling, Feature Selection, Multiheaded Attention Mechanism, Smart Grid Monitoring, Grid Stability Prediction, Class Imbalance, Energy Management

I. INTRODUCTION

Because of the huge rise in energy use and the use of these new energy sources and technologies, power networks rely on green energy sources and distributed energy systems in a very different way. Cutting-edge technologies like artificial intelligence (AI), the Internet of Things (IoT), and machine learning (ML) are used by modern smart systems to make the best use of energy exchange and consumption: You can now move from standard grids to smart grids, which use less energy, are more reliable, and leave less of a carbon footprint. Still, smart grids are getting more complicated, which makes them less stable and shorterens their useful life. Keeping such complicated systems running reliably and for good management depends on being able to predict their safety. In the context of the smart grid, stability means that it can handle problems and keep the balance of the networks that make, move, and distribute electricity. Changes in loads, the production of green energy, and exchanges between computers and the real world pose a major threat to the security of the grid. Accurate predictions of stability are needed to make strategic decisions and maybe even stop change from happening. At the moment, grid stability research methods are mostly built on set models that aren't very good at handling the dynamic and nonlinear parts of smart grids. So, we need to look into better prediction models that can handle complicated time series data, find hidden trends, and help us make decisions.

In recent years, deep learning models have become the best way to predict the security of a smart grid because they are specifically built for sequential data. It is well known that recurrent neural networks (RNNs), especially the Gated Recurrent Unit (GRU) and its variations, can show temporal relationships and non-linear trends in time-series data. Smart grid systems are appealing in GRUs because they are easier to set up and run than regular RNNs and Long Short-Term Memory (LSTM) networks. GRUs work well, but they could do even better if they used more methods to fix issues like wrong long-term predictions, uneven data, and choosing the right features. One of the biggest problems with forecast models for smart grids is that datasets are not balanced by nature. When models overfit to strong classes and fail on under-represented ones, unbalanced datasets could lead to biassed results. This problem only comes up when security goes through unusual but important changes. Using the Synthetic Minority Oversampling Technique (SMote) and other methods for data balancing has made it possible for minority groups to be better represented without changing how the data is spread out generally. Lessening the gap between classes would make it possible for more general and reliable prediction models.

The picking of traits is yet another important part of forecast models. Large files of smart grids may contain duplicate or unnecessary data that hurts the performance of the model. In feature selection tasks, particle swarm optimisation (PSO) and other bio-inspired optimisation methods have worked very well. PSO models group social activity to find the best feature subsets. This makes the model clearer and makes it more useful. When used with GRU networks, PSO helps choose the right input elements, which leads to more accurate and powerful results. The type of time series data that smart grids use means that they need complex systems to gather long-term ties and relevant links. Multiheaded attention methods, which were first used in transformer models, have shown a lot of promise in solving these issues. Multiheaded attention changes the importance of different time points in the input stream so that models can focus on important timing trends. This makes it easier for them to guess. Multiheaded attention and GRU networks work together to make a system that can find both short- and long-term links in data about grid stability.

This paper shows a mixed neural network structure that combines bio-inspired PSO planning, GRU networks, and Multiheaded attention methods to predict how stable smart grids will be: The method should take care of important problems in predictive modelling, such as data mismatch, feature selection, and long-term dependency modelling. The suggested model tries to make the most of the best parts of each part so that it can make better predictions than other methods. This study makes the field better in four ways. First, it lessens the bad effects of class mismatch in smart grid datasets by using Smote's data balance method. Second, it uses PSO to pick features so that the research can focus on the most important inputs. Third, it uses GRU networks and Multiheaded focus methods to make the model better at recording time dependencies. When everything is said and done, the framework's performance is carefully compared with that of simple models to show how well it can predict the smart grid are stable. The order of this paper is all over the place down here. In Part 2, important work is looked at with an eye towards current methods to smart grid security and what they can't do. The suggested method is explained in phase 3, which includes training techniques, feature selection, model building, and data preparation. In phase 4, the trial results are shown along with a review of how well the proposed model works compared to the commonly used methods. In phase 5, we talk about what the results mean, and in phase 6, we outline possible future study paths. Taking everything into account, our work fills a big need for advanced forecast models in the field of smart grid stability. Bio-inspired planning, GRU networks, and Multiheaded attention methods can all be used together to predict time series in settings that are complicated and change over time. Through this study, we hope to help make smart grid systems more reliable and long-lasting, so they can be used by more people and make money for a long time.

II. LITERATURE REVIEW

Finding out how safe smart grids are has become more important over the last few years as we've looked for more reliable and long-lasting energy sources. This literature study looks at the different methods that are used to predict the safety of smart grids today. It also looks at their flaws and opportunities for growth. To fix grid safety problems, people usually use set and random methods. Power flow analysis, transient stability assessment, and tiny signal stability analysis are some of the tools that are often used to keep an eye on and make sure grid stability. Even though these methods can teach us a lot, they can't fully capture the unexpected and irregular behaviour of modern smart grids because they are based on well-known rules and straight-forward assumptions. Also, these methods don't work very well with the huge amounts of time-series data that smart grids make. In the past few years, machine learning (ML) has shown a lot of promise for predicting how safe smart systems will be. Directed learning methods like ensemble models, decision trees, and support vector machines (SVMs) have been used to look into stable classification problems. These projects could help make more accurate predictions than current methods and find complex trends in data. For example, random forest models [8] have been used to predict how stable the power grid will be when green energy is used. These models can help you see trends that stay the same over time, but they're not great at handling data that changes over time. Deep learning is now a powerful tool for fixing problems thanks to smart grid time series forecast. A lot of people have used recurrent neural networks (RNNs) to copy the linear structure of grid stability data, especially Long Short-Term Memory (LSTM) networks. For example, LSTMs have done better than traditional models at predicting grid frequency safety at different load levels [9]. It's hard to write LSTMs, and their curves fade over time. This means that different systems, like Gated Recurrent Units (GRUs), need to be looked into.

GRUs interest people because they make it easy to duplicate time links with little computer work. The load safety of the smart grid has shown that GRUs can accurately predict it [10]. GRU models have some good points, but they might need to be improved in order to handle large amounts of data better and make more accurate long-term predictions. A common problem that makes smart grids less predictable is that datasets don't have enough rare but important stable events. One way to deal with this is to use oversampling methods, like SMote (synthetic minority oversampling technique). According to research, SMote could help machine learning systems better predict when the power grid might go down, which doesn't happen very often [11]. On the other hand, bad oversampling shows that we need better data balance methods because it can lead to overfitting and bad decisions. Picking the right features is important for getting accurate predictions and making forecast models easier to understand. People often use particle swarm optimisation (PSO), genetic methods (GA), and ant colony optimisation (ACO), all of which are based on living things, to do this. It was PSO that helped find the best sets of traits to use in time series forecast. When you combine PSO with deep learning models, it's easier to see what will happen in the future. This means that bio-inspired methods may be able to predict grid health. [12]. Sequence modelling has been changed a lot by Multiheaded focus systems that let models provide multiple data time stages with different levels of importance. At first, we found them in transformers, where they were It has been shown that attention processes help people understand everyday words pretty well. They can also be used to predict time series, as more study has shown. Multiheaded attention has been added to LSTM networks, making them more accurate and easier to use for estimating energy use in smart grids, for example [13]. Multiheaded attention hasn't been studied in depth yet, but combining it with GRU networks may lead to more accurate predictions of smart grid stability. Because they mix many methods, hybrid models are interesting because they let problems be solved without relying on just one. Predictions are more accurate and faster when mixed systems are used, which means using both deep learning models and bio-inspired planning. Putting PSO and LSTM models together should make green energy predictions better than using just one model [14]. This paper builds on earlier work by suggesting a mixed model that uses Multiheaded attention to make forecasts more accurate, PSO to pick features, and GRU to describe time.

Table 1. Summary of related research

Research Area	Technique	Advantage	Focus	Limitation
Traditional Methods for Grid Stability Prediction	Linear approximations	Limited scalability	Rule-based methods	Fails in dynamic conditions
Machine Learning and Smart Grids	Supervised algorithms	Higher accuracy	Handles complex patterns	Temporal dependency challenges
Deep Learning for Time-Series Prediction	LSTMs and GRUs	Captures temporal dependencies	Non-linear modeling	Computational complexity
Addressing Data Imbalance	SMOTE oversampling	Improves minority class performance	Better class representation	Risk of overfitting

Feature Selection Using Bio-Inspired Optimization	PSO and GA	Improves efficiency	Reduces dimensionality	Requires integration
Multiheaded Attention Mechanisms	Transformer-based models	Focuses on relevant patterns	Enhances interpretability	Underexplored with GRUs
Hybrid Models for Smart Grid Applications	Combines PSO and GRUs	Addresses individual limitations	Improves prediction accuracy	Requires seamless integration

Research Gaps and Contributions

While existing studies have made significant strides in smart grid stability prediction, several gaps remain. These include:

1. Limited exploration of GRU networks integrated with advanced attention mechanisms.
2. Inadequate handling of data imbalance in stability prediction datasets.
3. Suboptimal feature selection approaches that fail to leverage the full potential of bio-inspired optimization techniques.

This research addresses these gaps by proposing a novel hybrid framework that integrates PSO, GRU, and Multiheaded attention mechanisms. By addressing key challenges such as data imbalance, feature selection, and temporal dependency modelling, the proposed framework aims to advance the state-of-the-art in smart grid stability prediction.

III. DATASET

The 60,000-row synthetic dataset called the Smart Grid Stability dataset has 14 columns and is used to test grid stability in a wide range of situations shown in figure 1. Twelve numerical parameters are organised into three groups: delay features (Delay1, Delay2, Delay3, Delay4) show how the grid has changed over time; adapt features (adapt1, adapt2, adapt3, adapt4) show how to make changes; and power features (power1, power2, power3, power4) show metrics related to power. All of these factors affect how the grid works in real time, which is why they are so important for making predictions.

	tau1	tau2	tau3	tau4	p1	p2	p3
19247	1.908367	1.695836	3.059170	8.886750	3.609464	-0.870917	-0.818983
33302	7.362634	4.789448	6.638678	2.301323	2.985982	-0.594768	-1.501676
15148	4.188594	9.004552	4.129521	2.768301	4.437900	-1.462807	-1.271277
11891	2.734846	5.643794	6.893314	2.702091	3.674487	-0.750157	-1.545207
51801	6.431849	3.013871	5.272868	1.494029	4.337332	-0.534754	-1.891193

Figure 1. Dataset Sample

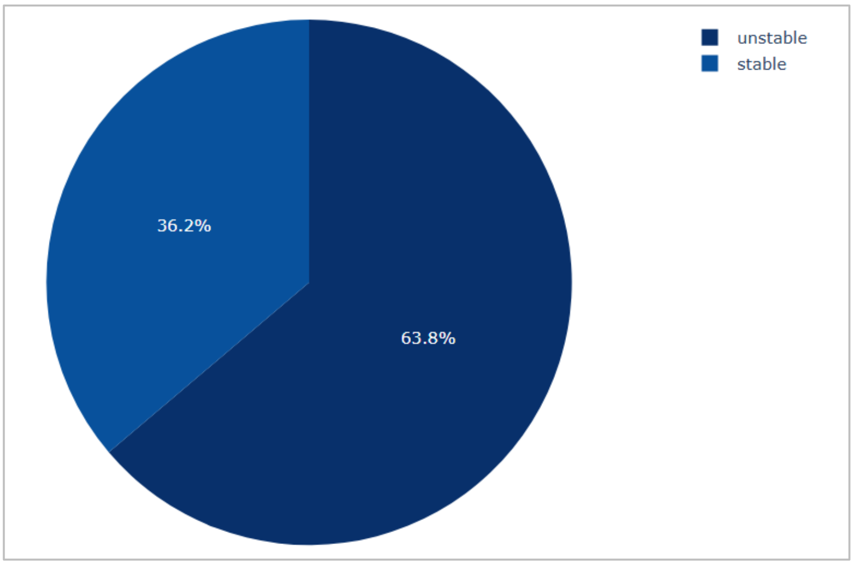

Figure 2. Data Distribution

Two target variables in the dataset are stab, a numerical value signifying grid stability, and stabf, a categorical variable with two classes—stable and unstable—often used for classification problems. The dataset shows some class imbalance; more unstable events than stable ones in figure 2.

IV. DATA PRE-PROCESSING AND BALANCING

Drop columns is the process of getting rid of unnecessary or duplicate data that has nothing to do with the prediction goal. For example, zero variance or strongly linked traits are taken out to lower the number of dimensions and the cost of processing.

Check for Missing Values One important first step is to look for missing numbers. If there are any, null or empty data points are found and dealt with. To keep the data clean, empty values can be filled in with statistical values like mean, median, or mode, or rows and columns with large amounts of missing data can be removed.

Data Normalization, particularly with Standard Scaler helps to make number features more consistent by setting their mean to 0 and their standard deviation to 1. This makes sure that features with different ranges or units don't have a big effect on the model while it's being trained, which improves speed and convergence.

Label encoding takes category factors and turns them into numbers. So that the model can handle classification data well, each group is given a unique numerical value. These pretreatment steps give you clean, consistent data that can be used for more advanced forecasting methods.

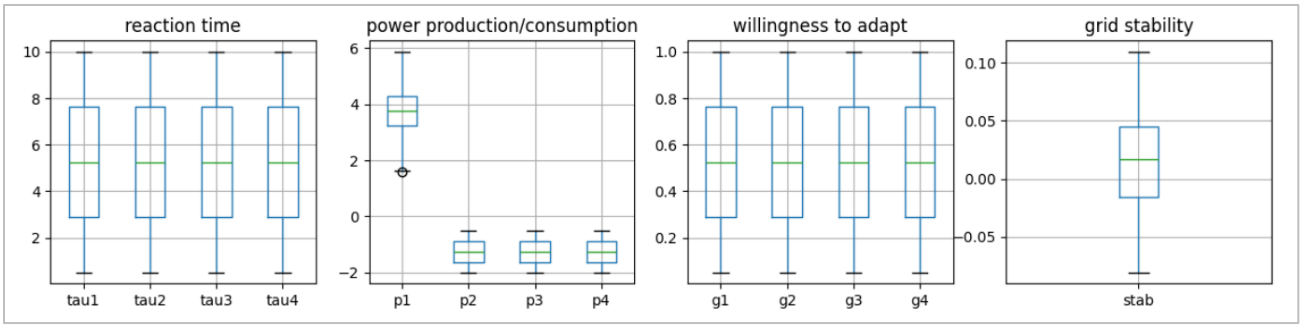

Figure 3. Data Visitation by Parameters

The Figure 3, Data Visitation by Parameters represents essential preprocessing steps for preparing the dataset to enhance the performance of the Bioinspired (PSO) + GRU Model.

```
stabf
unstable    0.638
stable      0.362
Name: proportion, dtype: float64
```

Figure 4. Sample Count in Percent (Before Data Balancing)

The Synthetic Minority Oversampling Technique (SMote) lets you take too many samples from the minority class to fix this problem shown in figure 4. This process makes fake cases to make the dataset more fair. This makes sure that both classes help the learning process equally. One could also pick the weighted loss functions or the sampling of the majority class. The balance that is being looked at affects how accurate the predictions are and how fair and right the model is in general.

V. FEATURE SELECTION

a. **LGBM Feature Importance**

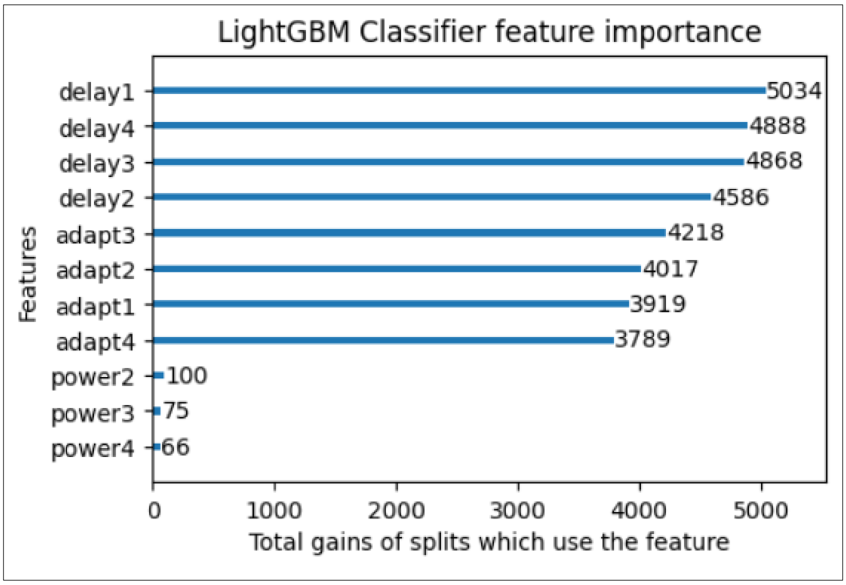

Figure 5: LightGBM Classifier Feature Importance

Figure 5 shows how the LightGBM Classifier calculated the importance of each feature by adding up the wins from all the breaks that used that feature. With total wins of more than 4,800, features like delay 1, delay 4, and delay 3 are definitely the best and make the model work better. The power characteristics (power2, power3, and power4) show small gains, which suggests they don't have much of an effect on the decision-making process. The adapt characteristics, on the other hand, show that they aren't that important either. LightGBM's ability to select features based on gradient boosting makes sure that the most important factors in the dataset are well understood. This image shows how important delay-based characteristics are for predicting smart grid stability, even though adapt-based features don't have much of an effect.

b. **RF Regressor permutation importance**

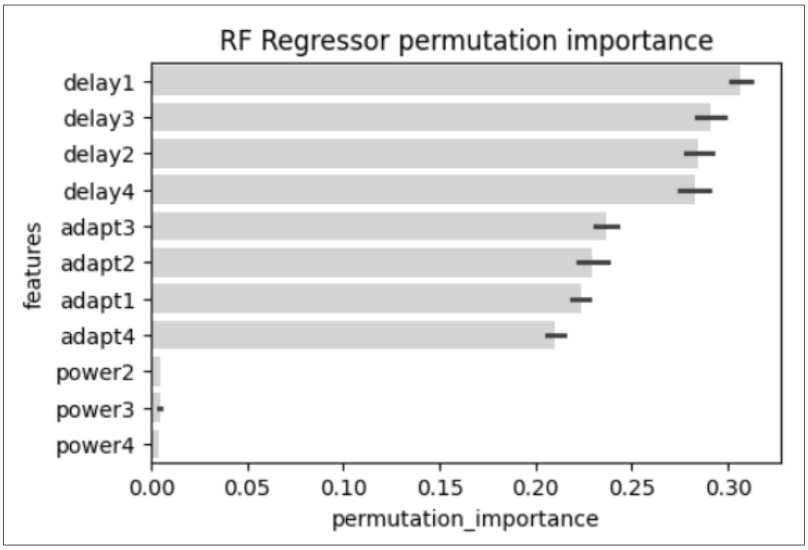

Figure 6: RF Regressor Permutation Importance

In Figure 6 Using the Random Forest Regressor, this graph shows how feature significance changes based on permutation significance. There is a strong connection between delay 1, delay 3, and delay 2 as shown by the fact that they have the highest permutation significance scores, which are around 0.3. The power features don't have much of an effect on the model, and neither do the adapt-based features (adapt3, adapt2, etc.). By switching around all the features and watching the model's performance go down, one can figure out the permutation significance, which is a strong sign of important predictions. This picture shows that while flexible features help a little, delay-based features clearly explain why the goal variable changes.

c. **Logistic Regression coefficients**

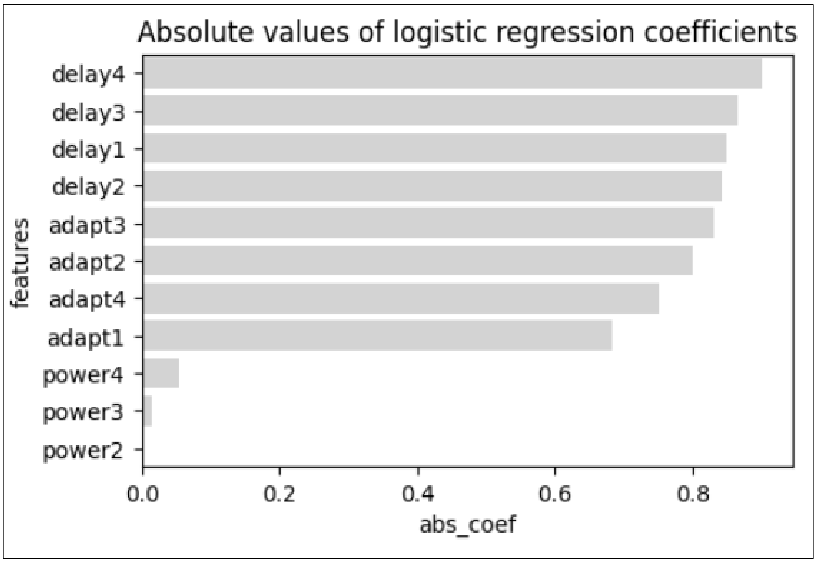

Figure 7: Absolute Values of Logistic Regression Coefficients

The exact values of the coefficients from a logistic regression model are shown in this bar chart. This shows how much each trait contributed to the estimate shown in figure 7. The delay traits are the most important; delay 4 and delay 3 have the highest values (about 0.8), which makes it clear that they have an effect on the model. Even though they are not as strong as the delay features, the adapt features also have important factors. The power-based traits, on the other hand (power2, power3, and power4) have low coefficients, which means they are not very good at predicting the future. In addition to showing how important delay-related characteristics are for making accurate predictions, the absolute coefficient values show how each feature is directly linked to the goal variable.

VI. PROPOSED HYBRID BIO-INSPIRED (PSO) + GRU MODEL

The building blocks of the Bioinspired (PSO) + GRU Model are shown in Figure 8. Its organised and flexible structure helps to guess how safe smart grids will be. The process starts when raw data from the file enters the Data Input Layer. The next layer is the preparation layer. To get the data ready for analysis, this is where important tasks like dimension reduction, data cleaning, normalisation, and SMote-based balance of the data are done. The feature selection (PSO) part of preprocessing makes sure that only the most important features are sent forward. Things move faster, and the machine doesn't have to do as much work.

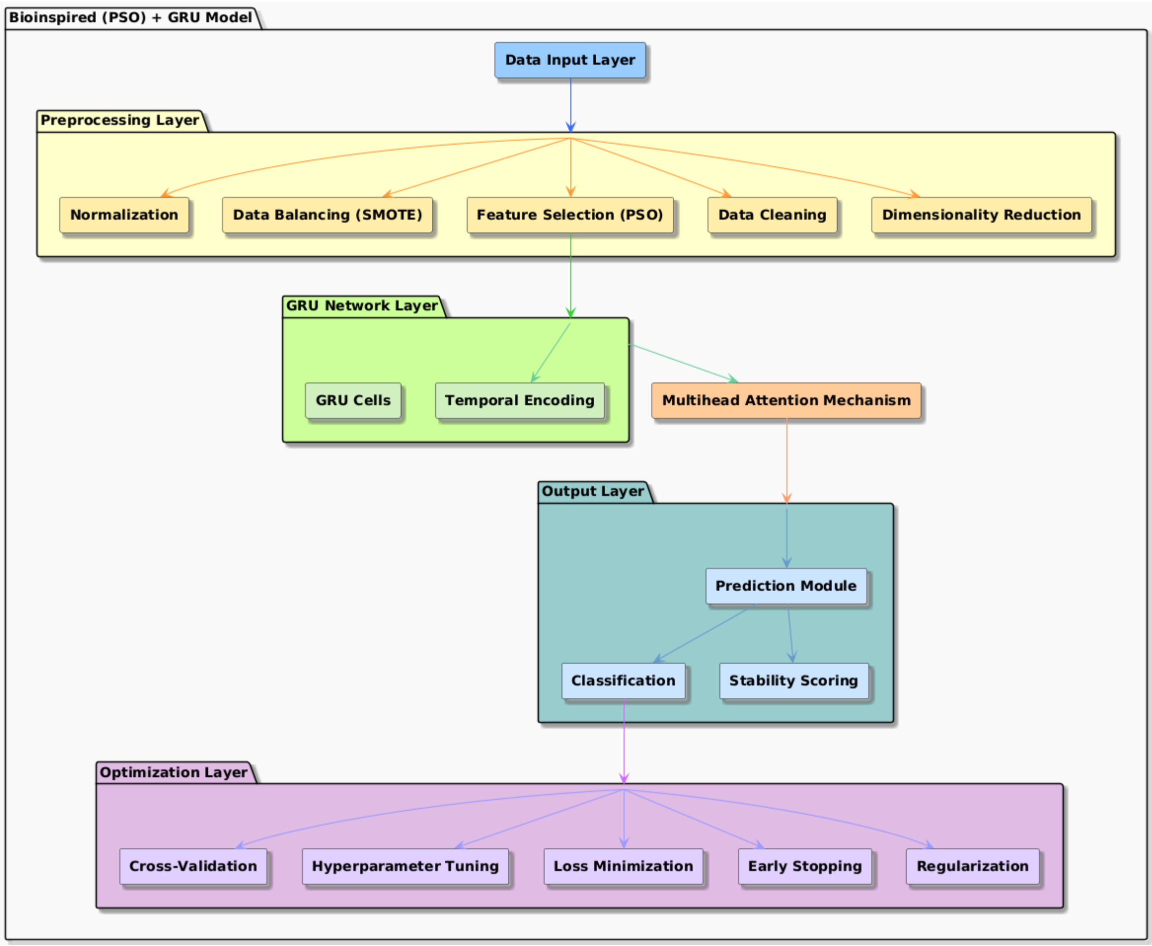

Figure 8. Architectural Framework of the Bioinspired (PSO) + GRU Model for Smart Grid Stability Prediction

The data that has already been handled is sent to the GRU Network Layer, which is made up of GRU cells and methods for encoding time to look for patterns and links in the input. The Multiheaded Attention Mechanism can make the model even more integrated by focussing on important time stages. This makes it easier to understand and leads to better performance. The Output Layer makes predictions, which are then sent to the Prediction Module. Depending on what is needed, it also ranks or groups steadiness. Finally, the Optimisation Layer makes changes to the model to make sure it works at its best. It uses cross-valuation, hyperparameter setting, loss minimisation, early stopping, and regularisation, among other things. This method makes sure that the smart grid's way of describing time data is strong, scalable, and effective.

PSO Feature Selection

```
Stopping search: maximum iterations reached --> 2
Selected Features: [ 0  5  6 11 12]
```

Figure 9. PSO Feature selction

The results show in figure 9 that the feature selection process stopped after the maximum number of cycles. This was most likely done with Particle Swarm Optimisation (PSO). The parts that were chosen stayed at [0, 5, 6, 11, 12]. These scores show which parts of the information are the most important. They help improve its performance by cutting down on the number of measurements, making the model more efficient, and keeping it from fitting too well. This step makes it easier for models like GRUs to make predictions by focussing on key factors like power load, voltage, and frequency and getting rid of variables that aren't important or are used more than once. These features are now being used as inputs for training models, which will speed up computations and make them more flexible.

GRU Integration

```
Model: "sequential_44"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 gru (GRU)                   (None, 5, 64)             12864

 gru_1 (GRU)                 (None, 32)                9408

 dense_90 (Dense)            (None, 1)                 33

=================================================================
Total params: 22305 (87.13 KB)
Trainable params: 22305 (87.13 KB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

Figure 10. GRU Integration

The sequence neural network on display is made up of two GRU layers and a thick output layer. The second GRU layer reduces the number of dimensions to (None, 32) and has 9,408 trainable parameters figure 10. The first GRU layer, on the other hand, has 64 units, looks at sequential input, and makes a shape of (None, 5, 64) with 12,864 trainable parameters. A final thick layer with one unit and 33 factors gives the outcome, which is good for tasks like regression or binary classification. The model is small and good at time-series tasks because it can extract strong temporal features. It has 22,305 trainable parameters, which is 87.13 KB, and there are no frozen layers.

Table 2. configuration of the Bioinspired (PSO) + GRU Model

Layer/Component	Configuration	Details
Data Input Layer	Input Shape: Depends on dataset dimensions	Accepts input features (e.g., power load, voltage, frequency) and passes them to preprocessing.
Preprocessing Layer		Handles data preparation tasks.
Normalization	MinMaxScaler or StandardScaler	Scales data to ensure uniformity and faster model convergence.
Data Cleaning	Null value handling, outlier detection	Removes missing or inconsistent values for cleaner input.
Dimensionality Reduction	PCA or Feature Truncation	Reduces feature space size to eliminate redundancy and improve computational efficiency.
Data Balancing (SMOTE)	Oversampling minority class	Ensures balanced data distribution for better generalization.
Feature Selection (PSO)	Number of particles: 50-100; Iterations: 50-200	Selects optimal features using Particle Swarm Optimization.

GRU Network Layer		Models temporal dependencies in sequential data.
GRU Cells	Number of GRU Layers: 2-3; Units per Layer: 64-128	Captures time-series patterns effectively.
Temporal Encoding	Attention Mechanism (optional); Dropout Rate: 0.2-0.5	Reduces overfitting and enhances learning of critical time steps.
Multiheaded Attention Mechanism	Heads: 4-8; Input Dim: Matches GRU output	Enhances focus on critical temporal dependencies, improving interpretability.
Output Layer		Produces predictions for grid stability.
Prediction Module	Fully Connected Layer; Activation: Softmax (classification) or Sigmoid (binary)	Maps encoded features to stability labels.
Classification	Output Shape: Number of stability classes	Classifies inputs into stability states (e.g., stable/unstable).
Stability Scoring	Regression or scoring model	Outputs a stability score to quantify risk levels.
Optimization Layer		Fine-tunes model performance through validation and tuning.
Cross-Validation	K-Fold (k=5 or 10)	Validates model on different subsets to improve generalization.
Hyperparameter Tuning	GridSearchCV or RandomizedSearchCV	Optimizes parameters such as learning rate, number of layers, and units per layer.
Loss Minimization	Loss Function: Binary Cross-Entropy or Categorical Cross-Entropy	Ensures better prediction by minimizing errors.
Early Stopping	Patience: 5-10 epochs	Stops training if validation performance does not improve.
Regularization	L1/L2 Regularization; Dropout: 0.2-0.5	Prevents overfitting by penalizing large weights and reducing over-reliance on specific neurons.

VII. RESULT ANALYSIS

The precision graph shows that the Bioinspired (PSO) + GRU Model works almost perfectly. The training accuracy quickly rises and stays close to 1.0 in the first few epochs, which shows that the model is learning from the data very well. The validation accuracy, which is very close to the training accuracy, shows that the model can work with data that it hasn't been able to work with before. This close match between training and evaluation accuracy shows that the model is not overfitting and is strong in figure 11. This conclusion is supported by the loss graph, which shows that in the first few epochs, both training and validation loss drop sharply before levelling off at zero.

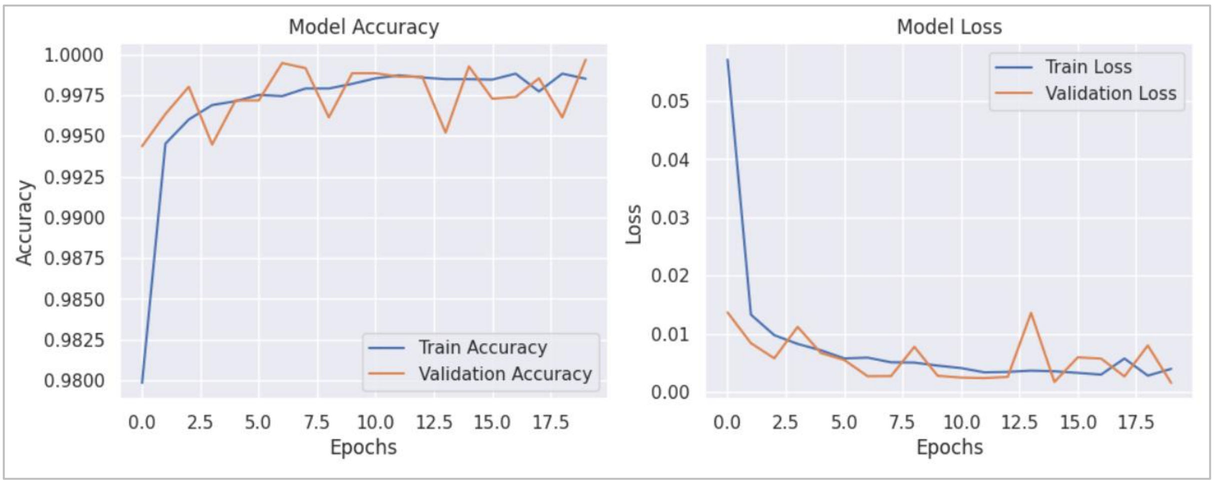

Figure 11. Accuracy and Loss Comparison Graph of Bioinspired (PSO) + GRU Model

As training loss goes down, it means that the model's predictions are getting more accurate. Also, the validation loss stays close to the training loss when few changes are expected in later epochs due to changes in the data. The model does a great job of handling time interactions and making reliable predictions, as shown by the small amount of loss and high accuracy for both datasets.

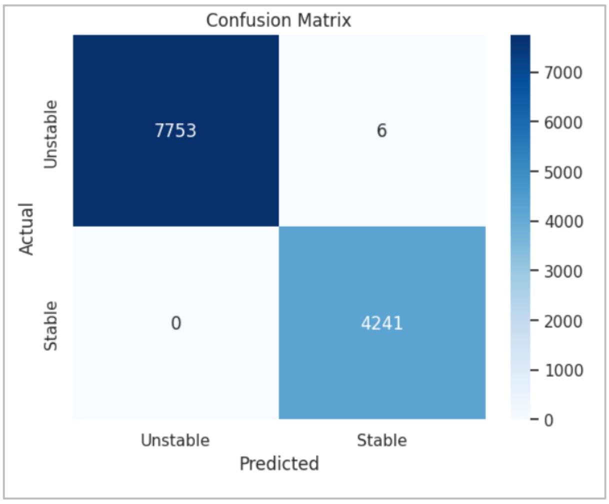

Figure 12. Confusion Matrix of Bioinspired (PSO) + GRU Model

The confusion matrix shows is figure 12, how well the Bioinspired (PSO) + GRU Model can predict the stability of the smart grid in both the "Stable" and "Unstable" categories. A lot of fake negatives and true positives were made by the model, which correctly called 753 cases "Unstable" and 4241 cases "Stable." Very few, just six "Stable" events were mistakenly marked as "Unstable." Also, there were very few cases of "Unstable" events being mistakenly labelled as "Stable." In terms of memory and accuracy, this means that the model is very close to being perfect. This almost perfect result shows how well the model can handle the complexity of the dataset and include data that hasn't been analysed yet. The low rate of wrong classification suggests that using GRU for modelling time and Particle Swarm Optimisation (PSO) for feature selection has led to a somewhat reliable way of estimating how reliable smart grids are. The results show that the model works well in real-life situations where quick and accurate predictions are needed.

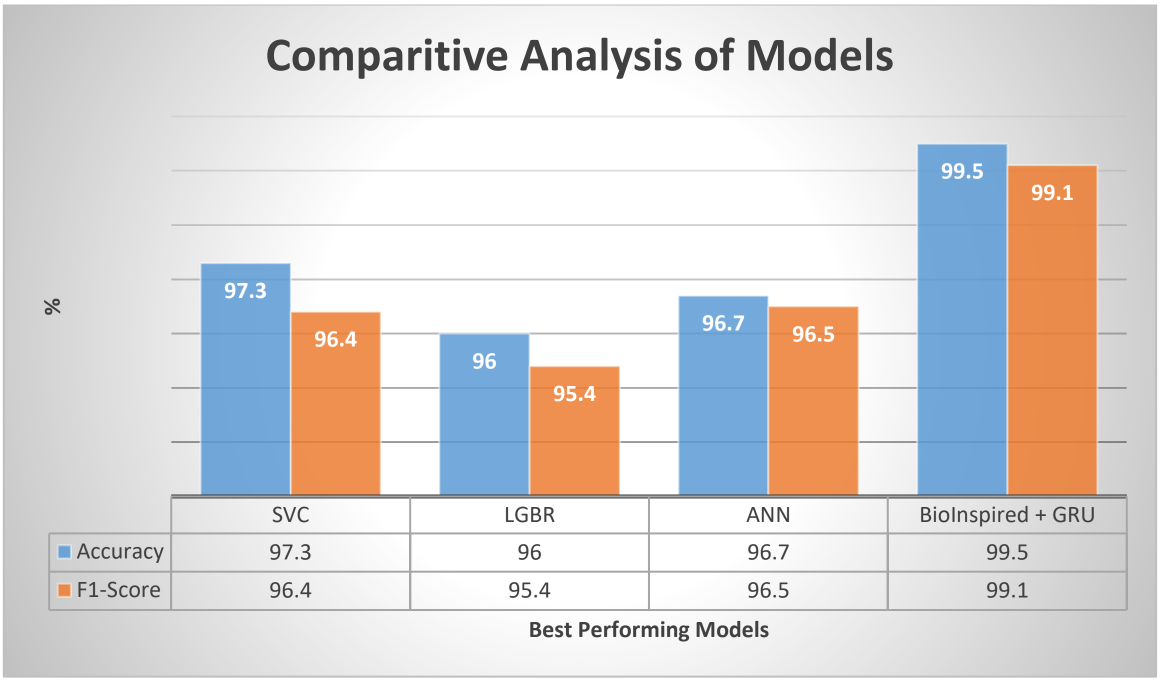

Figure 13. Comparative Performance Analysis of Models for Smart Grid Stability Prediction

We compare four models' performance here based on their F1-score and how well they work: SVC, LGBR, ANN, and the Bioinspired (PSO) + GRU Model in figure 13. Using this model, Bioinspired (PSO) + GRU, is the best way to predict and group stability. It's pretty good, with an accuracy of 99.5% and an F1-score of 99.1%. With a score of 96.5% on the F1 test, ANN comes in second. SVC and LGBR come in third and fourth, but their numbers are a bit lower. This paper shows how durable and effective it is to use both PSO for feature selection and GRU for time predictions. It does this by making both measures much better than other machine learning models, such as SVC and LGBR. The Bioinspired (PSO) + GRU Model works really well and is the best choice for important tasks where reliability and accuracy are very important, like predicting the safety of smart grids.

VIII. CONCLUSION

The Bioinspired (PSO) + GRU Model finally proves that it can handle datasets that aren't balanced and have complex temporal relationships. This makes it easier to guess how stable smart grids will be. The model reduces the number of variables, focusses on important characteristics, and finds repeated patterns in the data by using Gated Recurrent Units (GRUs) for time and Particle Swarm Optimisation (PSO) for feature selection. Its high accuracy (99.5%) and F1-score (99.1%) beat famous machine learning models like SVC, LGBR, and ANN in a study that compared them. The model is strong and reliable because it shows almost no false positives and false negatives in the confusion matrix. This is important for real-world use in smart grid systems. The model can generalise even more, as shown by the training and validation measures, since both the accuracy and loss graphs show steady growth over epochs without getting too well-fit. Some regularisation techniques that help the model last and be useful during training are dropping out and finishing early. When the GRU uses Multiheaded attention devices, it can focus on important time periods. The model can be used in real time because it is fast at computing and doesn't take up much space. It has 22,305 trainable parameters. This paper shows how useful it is to use bioinspired planning methods along with deep learning for high-dimensional, sequential data. Many fields, like banking, healthcare, and self-driving cars, depend on time trends and the importance of features, so the suggested method could also be used in those areas. If researchers want to get the best results in the future, they should look into better attention methods, more datasets, and mixed approaches. In conclusion, the Bioinspired (PSO) + GRU Model is a reliable and scalable way to make sure that smart grids are stable, which leads to smarter and more reliable energy systems.

REFERENCES

[1] K. Karthikumar, K. Karthik, K. Karunanithi, P. Chandrasekar, P. Sathyanathan, and S. V. J. Prakash, “SSA-RBFNN strategy for optimum framework for energy management in Grid-Connected smart grid infrastructure modeling,” *Mater. Today Proc.*, 2021.

[2] M. Massaoudi, H. Abu-Rub, S. S. Refaat, I. Chihi, and F. S. Oueslati, “Accurate Smart-Grid Stability Forecasting Based on Deep Learning: Point and Interval Estimation Method,” in *Proceedings of the 2021 IEEE Kansas Power and Energy Conference (KPEC)*, Manhattan, KS, USA, Apr. 19–20, 2021, pp. 1–6.

[3] M. Xia, H. Shao, X. Ma, and C. W. de Silva, “A Stacked GRU-RNN-based Approach for Predicting Renewable Energy and Electricity Load for Smart Grid Operation,” *IEEE Trans. Ind. Inform.*, vol. 17, pp. 7050–7059, 2021.

[4] B. Gao, X. Huang, J. Shi, Y. Tai, and J. Zhang, “Hourly forecasting of solar irradiance based on CEEMDAN and multi-strategy CNN-LSTM neural networks,” *Renew. Energy*, vol. 162, pp. 1665–1683, 2020.

[5] J. Li, D. Deng, J. Zhao, D. Cai, W. Hu, M. Zhang, and Q. Huang, “A novel hybrid short-term load forecasting method of smart grid using MLR and LSTM neural network,” *IEEE Trans. Ind. Inform.*, vol. 17, pp. 2443–2452, 2021.

[6] F. Mohammad and Y. C. Kim, “Energy load forecasting model based on deep neural networks for smart grids,” *Int. J. Syst. Assur. Eng. Manag.*, vol. 11, pp. 824–834, 2020.

[7] G. Capizzi, G. L. Sciuto, C. Napoli, and E. Tramontana, “Advanced and adaptive dispatch for smart grids by means of predictive models,” *IEEE Trans. Smart Grid*, vol. 9, pp. 6684–6691, 2018.

[8] P. R. Jeyaraj and E. R. S. Nadar, “Computer-assisted demand-side energy management in residential smart grid employing novel pooling deep learning algorithm,” *Int. J. Energy Res.*, vol. 45, pp. 7961–7973, 2021.

[9] B. Islam, Z. Baharudin, and P. Nallagownden, “Development of chaotically improved meta-heuristics and modified BP neural network-based model for electrical energy demand prediction in smart grid,” *Neural Comput. Appl.*, vol. 28, pp. 877–891, 2017.

[10] S. Gupta, F. Kazi, S. Wagh, and R. Kambli, “Neural network based early warning system for an emerging blackout in smart grid power networks,” in *Intelligent Distributed Computing*, Springer, Berlin/Heidelberg, Germany, 2015, pp. 173–183.

[11] B. Neupane, K. S. Perera, Z. Aung, and W. L. Woon, “Artificial neural network-based electricity price forecasting for smart grid deployment,” in *Proceedings of the 2012 International Conference on Computer Systems and Industrial Informatics*, Sharjah, UAE, Dec. 18–20, 2012, pp. 1–6.

[12] K. Verma and K. Niazi, “Determination of vulnerable machines for online transient security assessment in smart grid using artificial neural network,” in *Proceedings of the 2011 Annual IEEE India Conference*, Hyderabad, India, Dec. 16–18, 2011, pp. 1–5.

[13] M. Sakellariou and M. Ferentinou, “A study of slope stability prediction using neural networks,” *Geotech. Geol. Eng.*, vol. 23, pp. 419–445, 2005.

[14] J. V. Tu, “Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes,” *J. Clin. Epidemiol.*, vol. 49, pp. 1225–1231, 1996.

[15] S. Nijman, A. Leeuwenberg, I. Beekers, I. Verkouter, J. Jacobs, M. Bots, F. Asselbergs, K. Moons, and T. Debray, “Missing data is poorly handled and reported in prediction model studies using machine learning: A literature review,” *J. Clin. Epidemiol.*, vol. 142, pp. 218–229, 2022.

[16] S. Wang, S. Bi, and Y. J. A. Zhang, “Locational detection of the false data injection attack in a smart grid: A multilabel classification approach,” *IEEE Internet Things J.*, vol. 7, pp. 8218–8227, 2020.

[17] X. Niu, J. Li, J. Sun, and K. Tomsovic, “Dynamic detection of false data injection attack in smart grid using deep learning,” in *Proceedings of the 2019 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, Washington, DC, USA, Feb. 18–21, 2019, pp. 1–6.

[18] W. Sun, W. Lu, Q. Li, L. Chen, D. Mu, and X. Yuan, “WNN-LQE: Wavelet-neural-network-based link quality estimation for smart grid WSNs,” *IEEE Access*, vol. 5, pp. 12788–12797, 2017.

[19] S. Ungureanu, V. Ţopa, and A. Cziker, “Integrating the industrial consumer into smart grid by load curve forecasting using machine learning,” in *Proceedings of the 2019 8th International Conference on Modern Power Systems (MPS)*, Cluj-Napoca, Romania, May 21–23, 2019, pp. 1–9.

[20] M. Alamaniotis, “Synergism of deep neural network and ELM for smart very-short-term load forecasting,” in *Proceedings of the 2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, Bucharest, Romania, Sep. 29–Oct. 2, 2019, pp. 1–5.