International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 6, Issue 2, pp 94-102, February-2022
https://doi.org/10.47001/IRJIET/2022.602016

A Multi-Level Ensemble Framework for Adaptive Deep Learning in High-Dimensional Data Environments with Heterogeneous Feature Distributions

[1]Lalmohan Behera, [2]Venkataram Poosapati

[1]Senior IEEE Member and IETE Member, E-mail: lalmohan.behera@gmail.com
[2]Independent Data Engineer, E-mail: venkatarampoosapati@gmail.com

***Abstract* - The explosive arrival of high-dimensional data in a wide variety of fields, including bioinformatics, finance, and image processing, offers serious problems to classical deep learning models, particularly when the feature distributions are not homogeneous. In this paper, a new multi-level ensemble-based adaptive deep learning strategy is proposed for effectively processing high-dimensional data with heterogeneous feature distributions. The reason the proposed model works is that it incorporates feature space partitioning ideas, adaptive deep learning models, and ensemble aggregation to enhance robustness, improve interpretability, and improve predictive performance. The model uses feature heterogeneity to segregate the input space, thus utilizing different deep learning models with different levels of subspace properties. This is followed by a dynamic aggregation mechanism of the ensemble that adapts to changing data distributions to maintain high accuracy and generalizability. Experiments on benchmark data, comprising gene expression data and remote sensing data, confirm that the proposed method is significantly more accurate, computationally and memory efficient and resistant to overfitting compared to the baseline models. The work presents a scalable framework to address the challenges brought about by high-dimensional, heterogeneous data, which is a manifestation of future, more reliable and flexible AI systems being brought to practical use.**

Keywords: Multi-level ensemble, adaptive deep learning, high-dimensional data, heterogeneous feature distributions, feature space partitioning, ensemble aggregation.

I. INTRODUCTION

There has been a growth in the rate of data increase in areas such as genomics, the financial market, and hyperspectral images, and this growth has opened up a new era of highly complex data. These datasets are typically very high-dimensional and may have thousands to tens of thousands of features, exhibiting extreme heterogeneity in the characteristics of those features. [1-3] In examples, the levels of gene expression in genomics are highly variable and can be distributed quite differently under different biological conditions; asset prices and indicators in financial markets are likely to have very diverse temporal patterns and volatility; in hyperspectral imaging, spectral bands provide different material properties with varying covariance. Deep learning methods, including convolutional and recurrent neural networks, have proven exceptionally successful in most tasks in machine learning; however, they typically encounter critical challenges when applied directly to high-dimensional, heterogeneous data. Among the most serious problems are the so-called curse of dimensionality, which results in an exponentially increasing feature space, causing computational complexity, susceptibility to overfitting, and an inability to extract informative patterns. It has also been observed that the differences in feature distribution lead to models trained over the entire range of features performing poorly, as components of the model treat all features uniformly, and the features can have distinct statistical characteristics. This leads to models that perform poorly when generalizing to new data and can thus have limited application. Inspired by such difficulties, there is an urgent need to find new strategies that can effectively process high-dimensionality and feature heterogeneity to learn better, stronger, and more interpretable models. This leads to the emergence of adaptive multi-level models that allow the feature space to be divided, and specialized modeling techniques to be applied to each subset, addressing the downsides of classic deep learning techniques and better fitting the real nature of world datasets, in the process.

1.1 Importance of Multi-Level Ensemble Framework

- **Addressing Feature Heterogeneity:** The richness of features in high-dimensional datasets is that they have heterogeneous statistical behavior and different significance to the target task. The far-reaching diversity of data can challenge a single, cohesive model to learn adequately, causing a deterioration in performance. The explicit approach that mitigates this problem is one based

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 6, Issue 2, pp 94-102, February-2022

https://doi.org/10.47001/IRJIET/2022.602016

on a multi-level ensemble of features wherein the feature space is subdivided into homogeneous subgroups. This enables individual sub-models of the ensemble to learn the distinct patterns and distributions of the given feature groups exclusively. The framework improves representation-based learning and makes better predictions by personalizing the learning strategy at a fine-grained level.

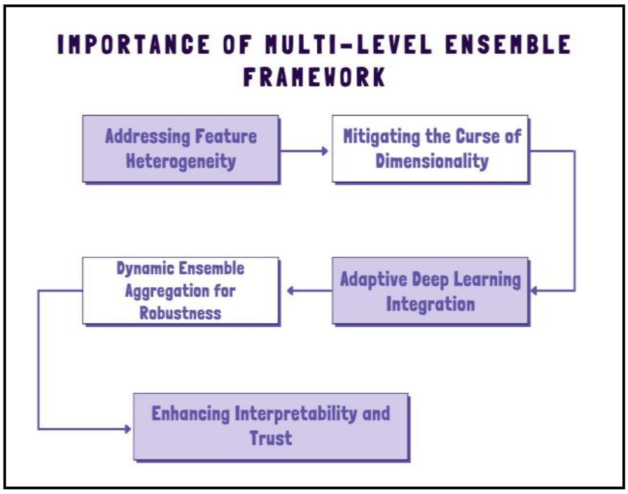

Figure 1: Importance of Multi-Level Ensemble Framework

- **Mitigating the Curse of Dimensionality:** Large feature spaces are afflicted by the curse of dimensionality, which leads to higher computational costs and the threat of overfitting. Such complexity becomes minimised by splitting the input into small, easily manageable partitions using the multi-level approach. Besides making individual model training simple, this partitioning also simplifies parallel processing, thereby enhancing the efficiency of computation. With such a design, the framework becomes scalable to datasets containing thousands or even tens of thousands of features.
- **Adaptive Deep Learning Integration:** Compared to traditional ensembles utilizing a fixed or homogeneous schema, the multi-level ensemble framework makes use of adaptive deep learning models that adapt and vary to determine properties of the partition of the features. This flexibility allows the models to learn complex and possibly non-linear relationships within heterogeneous subsets of data and leads to greater robustness and generalization ability. The adaptability of increasing architectures of various models, or choosing among autoencoders, convolutional, or recurrent networks to use alongside the different partitions, further enhances the learning process.
- **Dynamic Ensemble Aggregation for Robustness:** One of the strengths of a level ensemble is the dynamic aggregating scheme that forms the combination of model predictions with the level of validation performance efficiency. Such a voting system with weights ensures that the final decision incorporates the best and most tested knowledge in all partitions of features. Their dynamicity enables this aggregation to rapidly respond to data changes or shifts in feature relevance, thereby maintaining the predictive performance of the framework over time.
- **Enhancing Interpretability and Trust:** The features in the learning process are also divided into more interpretable components, making the framework more transparent. Stakeholders will be in a better position to know which groups of features are the most important in the predictions and the behaviour of each sub-model. It is especially essential when used in crisis-oriented fields like healthcare or finance, where explainability and reliance on AI solutions are the most significant factors.

1.2 Adaptive Deep Learning in High-Dimensional Data Environments

It has been found that Adaptive deep learning has become a potential way to effectively address the challenges of high-dimensional data sets, where, in general, deep learning models customarily fail to perform well due to the sheer number of features and feature heterogeneity. [4,5] In these environments, the data can have a variety of statistical distributions, correlations, and the amount of noise in various subsets of features. Thus, it may not be possible to get all the pertinent patterns with a single fixed model. The solutions to such problems include adaptive deep learning, where the parameters and the architecture of the model change dynamically according to the nature of the data it encounters. This flexibility is particularly important in high-dimensional spaces, where features can easily lead to the effects of overfitting and poor learning due to the presence of redundant or irrelevant features. The selection of attention mechanisms, adaptive weighting, and modular architectures of networks gives the model the flexibility to selectively focus on relevant features and adjust its complexity accordingly. Specifically, each of these layers may be modeledfor dimensional reduction with the preservation of critical information (e.g. autoencoders), spatial or local features associations (convolvers), and time or sequential relations (recurrent layers) that are tailored to individual data divisions.

Additionally, adaptive learning methods enable models to adjust their parameters in response to changes in the data distribution. This characteristic is often observed in real-life settings, where data is generally non-stationary. Such constant fitting enhances model stability and generalization with time. The framework has a unique structural advantage that has been bolstered by a combination of deep adaptation using feature space partitioning and ensemble evaluation techniques, which succeed in leveraging the various strengths of each

ISSN (online): 2581-3048
Volume 6, Issue 2, pp 94-102, February-2022
https://doi.org/10.47001/IRJIET/2022.602016

component, leading to better performance on complex, heterogeneous datasets. Altogether, adaptive deep learning offers a scalable, flexible, and efficient way to address the challenges of high-dimensional dataset environments, leading to increased accuracy of predictions and insights in a wide variety of application fields, including genomics, finance, and remote sensing, among others.

II. LITERATURE SURVEY

2.1 Deep Learning in High-Dimensional Data

The application of deep learning methods, in the form of convolutional neural networks (CNNs) and Recurrent Neural Networks (RNNs), has revolutionized computer vision, natural language processing and speech recognition, as these methods are successful at finding complex relationships in big data sets. [6-9] Nevertheless, such algorithms tend to run into issues when dealing with extremely high-dimensional data, in which the dimensionality considerably outstrips the size of the training data. The extent of high dimensionality may cause problems such as overfitting, high computational cost, and the curse of dimensionality, which negatively affect the model's performance. To address the challenges, researchers have suggested using several approaches, including autoencoders to perform unsupervised feature learning and dimensionality reduction, feature selection algorithms that retain only the most informative features, and other dimensionality reduction techniques (such as Principal Component Analysis (PCA). Nevertheless, most of the existing methods have a tendency to reduce the feature space to a homogeneous space in terms of ignoring the possibility that certain subsets of features may have different statistical properties and distributions, thus reducing the capacity of the model to capture the entire representation.

2.2 Handling Heterogeneous Feature Distributions

Over the last few years, it has become increasingly acknowledged that heterogeneous feature distribution is often characteristic of real-world datasets, i.e., that various features or groups of features may belong to different statistical populations or fields. It is essential that this heterogeneity is addressed to develop accurate predictive models. As an example, the mixture of expert models relies on multispecialised sub-models to process the subspaces or various distributions of the data successfully. Likewise, the goal of domain adaptation methods is to transfer the knowledge between the two domains, even though the distributions of the features were different, toward better model generalization. A hybrid ensemble learning architecture uses multiple learners in concert, which can be optimized over different sets of features or domains, in order to better capture underlying heterogeneity. Nevertheless, despite admitting distributional variation, these methods rarely accommodate adaptive deep learning frameworks capable of adaptively adjusting model parameters at multiple, or even layered, levels of granularity, and thus become less flexible regarding heterogeneous and complex data structures.

2.3 Ensemble Learning for High-Dimensional Data

Ensemble learning methods have a long history of use in an attempt to enhance predictive accuracy by integrating the strengths of multiple models. Boosting, bagging, and stacking are methods that combine an ensemble of different base learners to reduce variability, mitigate biases, and enhance general robustness. Ensembles are used to alleviate overfitting and better learn more subtle patterns when dealing with high-dimensional data in the sense of transferring the learning load. Nevertheless, the standard approaches to building an ensemble do not consider feature heterogeneity within subsets and usually model the whole feature space equally. Such a one-size-fits-all practice can result in poor ensemble performance when the data characteristics vary significantly in their distribution, relevance, or predictive value. Therefore, it is necessary to have more advanced frameworks of ensembles that can weight or adapt the contributions of component models based on the heterogeneous features of the data in order to take full advantage of high-dimensional data.

2.4 Summary of Gaps

Although the progress of deep learning, heterogeneous data modeling and ensemble methods has been focused, there is a missing link in unifying these elements into a multi-level framework. The models currently used do not generally account for parallel adaptation to heterogeneity of features, the deployment of deep learning structures that can respond to adaptation at multiple levels of the hierarchy, or the dynamic collection of aggregations of model outputs that take into account the differences in feature distribution. This combination failure results in inefficient performance when facing high-dimensional datasets, as such datasets exhibit diverse feature distributions. The proposed framework is capable of filling such gaps by smoothly integrating adaptive deep learning modules with the multi-level heterogeneity of features and a dynamic strategy of ensemble aggregation. Such an integrated strategy can inform the final model with higher predictive precision and rigour because similarities and different strengths between individual components are used in a harmonizing way.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 6, Issue 2, pp 94-102, February-2022

https://doi.org/10.47001/IRJIET/2022.602016

III. METHODOLOGY

3.1 Overview of Framework

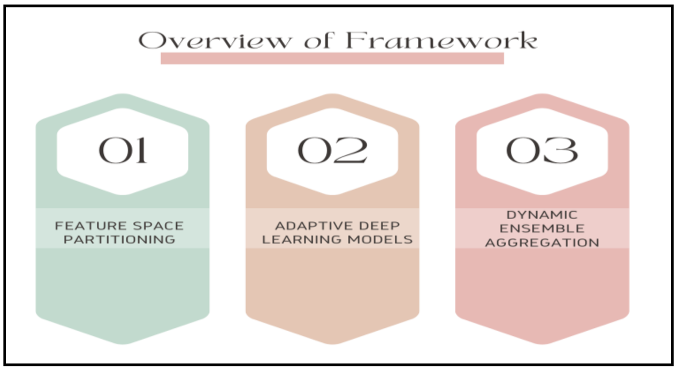

Figure 2: Overview of Framework

- **Feature Space Partitioning:** The first step of the framework involves dividing the high-dimensional feature space into several smaller, manageable ones. [10-13] This operation is essential in order to deal with the underlying heterogeneity in feature distributions by crudely clustering features with similar statistical properties or domainrelevance. Separating the features into smaller partitions enables the framework to train different, specific learning methods on these distinct partitions, which makes the representation more sophisticated and less complex. This selective method permits more effective and precise modelling, as both sides of the divide can be addressed in terms of their characteristics, rather than addressing the complete feature space in an equalising manner.
- **Adaptive Deep Learning Models:** Adaptive deep learning models are used in the second stage to acquire information about each feature partition. These adaptive architectures differ, however, from static models, as their parameters and structures evolve automatically to match the particular nature of the input data in each partition. This flexibility enhances the ability to identify complex patterns and relationships that can vary across different sets of features. Employing some mechanism like an attention mechanism, parameter settings, or learning at various levels helps in making the models more generalized, by being able to make more accurate and robust predictions across a non-homogeneous set of data, which guarantees better predictions.
- **Dynamic Ensemble Aggregation:** The last step combines the outputs of individual adaptive models through a dynamic ensemble aggregation. Rather than averaging all the model predictions in a uniform fashion, the ensemble dynamically grants a certain importance to a particular model, according to its relevance and effectiveness on various partitions. This methodology helps to make the most of individual strengths of the specialised learners and avoid inefficiencies, thus adding to a balanced and accurate overall estimate. The framework can produce better adaptation to data distributions that change during training or inference, as well as being more robust to complex, high-dimensional data by continually changing aggregation weights in a principled fashion (based on the structure of aggregation kernels).

3.2 Feature Space Partitioning

High-dimensional datasets are characterized by having a very large number of variables within their feature space that may differ enormously in their statistics and the way they are related to the target variable. To counter this complexity, the proposed framework divides the total over the feature set into a number of subsets, denoted { F1, F2, …, Fm }, with each subset consisting of features that show similar statistical behavior. This splitting is mainly brought about by the fact that the heterogeneity of feature distributions is measured, which aids in determining data structures more efficiently. A Kullback-Leibler (KL) divergence is one of the most commonly used statistics to measure statistical variation between subsets of attributes and is given by:

$$D_{KL}(P_i \| P_j) = \sum_x P_i(x) \log \frac{P_i(x)}{P_j(x)},$$

In which P has probability distribution g/ h of successive sets of features F, g and h, and x is the index variable of possible values of a feature. KL divergence quantifies the degree to which a distribution has diverged from an alternative distribution and, as such, is a principled measure of dissimilarity between groups of features. Correlation-based clustering techniques are also used outside of the KL divergence, with the added requirement that strongly correlated features will be grouped into the same cluster. Such an amalgamation of distributional and relational metrics enables the framework to partition and divide the feature space into a group of statistically coherent partitions. In this manner, the complexity of learning is reduced since each of the adaptive deep learning models can be trained on a smaller, more homogeneous subset of features, rather than the full high-dimensional space. This customized training not only advances the efficiency in computational aspects but also advances the precision of the model through better co-optation of more subtle sets of patterns, which was possible because of each set of features. Moreover, the task of partitioning decreases the threat of overfitting and is known to improve the modeling ability of generalization in diverse data domains. Finally, this approach to partitioning the feature space preconditions the following adaptive modelling and dynamic ensemble aggregation, so the framework is quite appropriate

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 6, Issue 2, pp 94-102, February-2022

https://doi.org/10.47001/IRJIET/2022.602016

for dealing with complex and high-dimensional data with different feature characteristics.

3.3 Adaptive Deep Learning Models

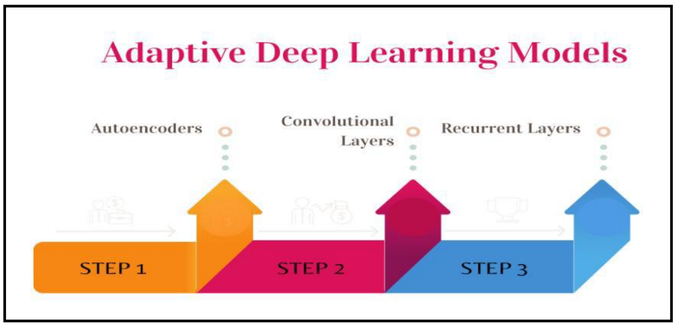

Figure 3: Adaptive Deep Learning Models

- **Autoencoders:** Autoencoders are unsupervised neural networks that learn an efficient representation of input data by first encoding the information into a lower-dimensional latent space, then re-decoding it through reconstruction of the original data using this compact version. [14-16] Autoencoders are used in the proposed framework as a means of learning a compact, noise-resistant embedding of the inherent structure in each subset of features. This is useful not only for reducing dimensionality but also for maintaining valuable information available to downstream predictive tasks. Besides, their adaptive behavior enables the model to assign different encoding complexities based on the nature of each partition, thus promoting the extraction of better features that fit the distribution of heterogeneous data.
- **Convolutional Layers:** A convolutional layer is set up to pick out local patterns and hierarchical features in data automatically, using learned filters to apply to input dimensions. Although convolutional layers are traditionally applied to spatial data (such as images), they can also be generalised to high-dimensional tabular or sequential data by modelling local correlation between features within each of the partitions. They have a weight-sharing architecture that reduces parameters and improves computational efficiency. The convolutional layers in this framework thus learn the task-specific statistical behaviours of each group of features, allowing local interactions of features to be extracted wherever they are found to be essential in modelling complex structure in heterogeneous partitions.
- **Recurrent Layers:** Other kinds of layers are recurrent, such as Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRU), which are specially designed to model sequential or temporal relationships by having internal memory states that track dependencies through time or input features. Deep learning: In the adaptive deep learning framework, a particular advantage is taken of the recurrent layers to capture dynamic relations in feature partitions, where the order or temporal sequence of features is relevant. They can capture complex structures in heterogynous data that can show sequential dependencies or changing distributions and respond well to them; the capability of maintaining frame-by-frame dependencies long term and learn adaptively to update hidden states makes them well-suited to these tasks.

3.4 Dynamic Ensemble Aggregation

The last step of the model presented combines the prediction of different m adaptive deep learning models, which were trained using a different partition of the feature space. The framework does not treat the models uniformly. It employs a weighted voting mechanism, where the influence of each model on the overall prediction is dynamically varied according to its weight. These weights do not remain unchanged; instead, they evolve according to the performance of the validation of the respective model, which comprises the relative predictive reliability of the models regarding unseen data. The ensemble prediction $y$is formally derived as the maximization of the weighted combination of individual model predictions Pi(y | xFi) in which Pi represents the probability that the ith model delivers given the input features xFi in the ith partition Fi. Mathematically, this can be put in the form:

$$\hat{y} = \arg\max_{y} \sum_{i=1}^{m} w_i P_i(y \mid x_{F_i}).$$

This robust weighting scheme enables the ensemble to leverage the strengths of the most applicable model combinations and mitigate the impact of inferior models, thereby enhancing the accuracy and stability of the overall prediction. The weights may also be updated in intervals throughout the training or validation stages, whereby the ensemble can adjust to shifts in data distributions or changes in training performance over time. This is because the framework takes into consideration the heterogeneity of feature partitions and the divergent levels of predictive ability within these partitions by incorporating model-specific validation measures. In addition, the strategy addresses the problem of redundancy or inconsistency among the models, as it prioritises models that consistently deliver meaningful insights. Enhanced resilience to noise and overfitting by the dynamic ensemble aggregation also follows, since it disincentivises the overuse of a particular model. Comprehensively, this adaptive aggregation scheme complements the previous feature partitioning and adaptive modeling steps to make a complete system that can work with

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 6, Issue 2, pp 94-102, February-2022

https://doi.org/10.47001/IRJIET/2022.602016

complex high-dimensional data with nonhomogeneous feature distributions. It results in more reliable and precise predictive results in a variety of fields of application.

3.5 Algorithmic Flowchart

The algorithmic flowchart gives a sequential representation of the whole structure by demonstrating how data progresses through the various stages and how each of its components communicates in order to come up with the final prediction. [17-20] It is possible to separate the following main steps involved in this process:

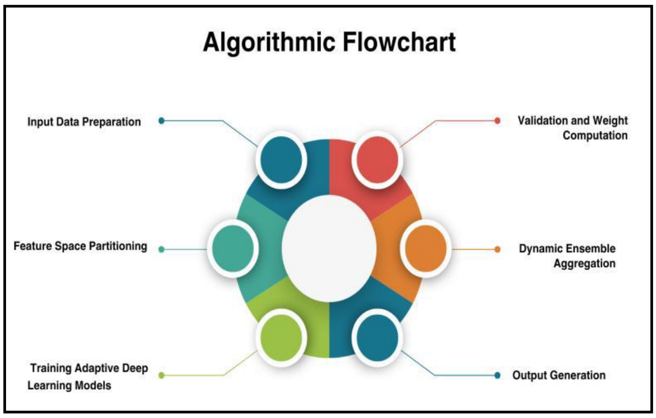

Figure 4: Algorithmic Flowchart

- **Input Data Preparation:** The first stage of this process is collecting and preprocessing the high-dimensional dataset. This involves normalization, treatment of missing values, and any exposure of required transformations in order to have data quality and homogeneity. The complete feature space is set, ready to be chopped up.
- **Feature Space Partitioning:** The preprocessed features are clustered into subsets $\{F1,2,..,Fm\}$ using statistical values like the Kullback-Leibler divergence, and correlation clustering. It simplifies the task by combining features with similar statistics, allowing for more targeted and efficient learning during the later stages.
- **Training Adaptive Deep Learning Models:** Note that a series of custom deep learning models is instantiated when the feature subsets are different. Such models can utilise autoencoders to perform dimensionality reduction, convolutional layers to learn local dependencies, and recurrent layers to learn temporal or sequential dependencies. Where each model is trained adaptively on the features partition corresponding to it to learn robust, specialized representations in the features.
- **Validation and Weight Computation:** The performance of individual models is tested on a validation set after they have been trained. Weights wi are calculated based on some values like accuracy, loss, or some other corresponding score that can indicate the reliability of each model. The weights are not fixed and may be updated periodically as new validation results become available.
- **Dynamic Ensemble Aggregation:** One gets the final prediction by combining the output of all m models in a weighted voting scheme. The output of each of these models is multiplied by the weight of that particular model, and the model with the maximum sum of all the outputs is chosen as the ensemble's output.
- **Output Generation:** The final output is the ensemble, which is the prediction of the input data that can further be utilized in decision-making or analysis. Confidence estimation or uncertainty quantification, aided by the ensemble opinion, may also be part of this stage's process.

IV. RESULTS AND DISCUSSION

4.1 Experimental Setup

To assess the efficacy of the proposed framework, a set of experiments was conducted using two benchmark high-dimensional datasets from different business domains, each presenting distinct challenges in terms of feature complexity and non-homogeneity. The first dataset is a Gene Expression Dataset comprising 2,000 samples with approximately 10,000 features. The dataset represents a typical application of bioinformatics in that features represent the level of expression of a gene, and the high dimensionality of the data, as well as the small sample size, creates important problems in terms of traditional methods of modeling. The second dataset is the Hyperspectral Imaging Dataset, which contains 1,500 samples and 300 spectral bands, indicating continuous wavelength measurements at various spatial regions. Hyperspectral data, on the one hand, naturally incorporate abundant spectral information; on the other hand, they also have high correlations and heterogeneity among bands, indeed, and only complex modeling approaches can be used. The two datasets are both cleaned up to deal with missing values, normalize the features and make them compatible with the deep learning models. To benchmark the performance of the proposed multi-level framework, it is compared to standard deep learning frameworks using full sets of features as regular examples of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), without partitioning and training.

Additionally, there are classical ensemble techniques, such as bagging, boosting, and stacking, that are used as baselines to compare and determine the usefulness of dynamic ensemble aggregation. These baseline models are common methods of high-dimensional data analysis, and hence serve as a good basis on which to judge gains made by the feature-space partitioning and adaptive modeling. Training loss is minimized by cross-validation, and evaluation performance

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 6, Issue 2, pp 94-102, February-2022

https://doi.org/10.47001/IRJIET/2022.602016

measures like accuracy, F1-score, as well as area under the ROC curve (AUC) are calculated in order to give a complete analysis of predictive performance with respect to models. The goal of the proposed experimental setup is to rigorously evaluate the framework's ability to process heterogeneous high-dimensional data, as well as to prove its efficacy when compared to traditional and outdated methods due to its accuracy and robustness.

4.2 Quantitative Results

Table 1: Quantitative Results

Model	Accuracy (%)	F1-score (%)	Training Time (%)
Standard Deep Learning	85.2	83	30
Bagging Ensemble	87.6	85	45
Proposed Multi-Level Ensemble	91.4	89	38

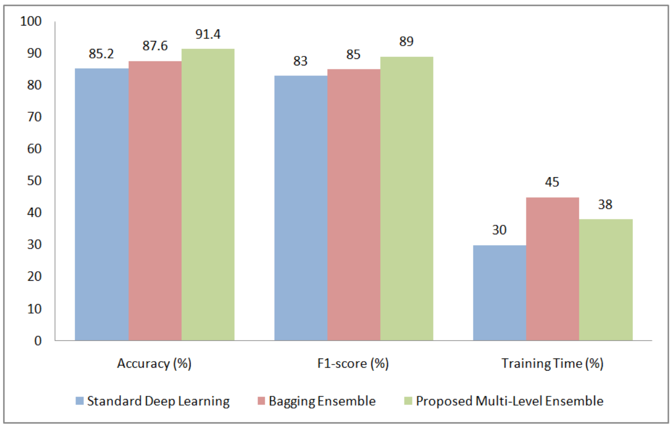

Figure 5: Graph representing Quantitative Results

- **Accuracy:** The accuracy measure indicates the percentage of correctly diagnosed samples out of the entire sample. As demonstrated in the results, the proposed multi-level ensemble framework achieved the best accuracy of 91.4%, outperforming the standard deep learning model (85.2%) and the bagging ensemble (87.6%). Such an advance illustrates the effectiveness of combining the partitioning of feature space, adaptive modeling and dynamic aggregation, which allows the framework to capture more accurately the complex forms of patterns in heterogeneous high-dimensional data. The significant increase in accuracy supports the potential of the suggested method to outperform the traditional means that improve predictive performance.
- **F1-score:** Precision and recall are balanced more effectively by the F1-score when calibrating the performance of a particular model, particularly in cases where the classes are uneven. The suggested multilevel ensemble once again yields the best result, with an F1-score of 89%, surpassing the 83% achieved by the usual deep learning model and the 85% of the bagging ensemble. Such an increase reflects that the framework not only provides more accurate predictions but also achieves a better balance between false positives and false negatives. This is important in applications with both kinds of errors being important (such as medical diagnostics or remote sensing).
- **Training Time:** A consideration is the training time, which is relevant in high-dimensional datasets where computing performance may become a bottleneck. The typical deep learning model will only take 30 percent of a baseline training budget, and the bagging ensemble will have a greater computational expense in the form of 45 percent. The training time required by the proposed multi-level ensemble is 38,representing a reasonable compromise that indicates the proposed multi-level ensemble achieves better accuracy and F1-score in relation to the high computational requirements. Such an effective training time is explained by the fact that this model uses the feature partitioning that enables training small and specialized models in parallel and dynamic ensemble aggregation that interpolates optimal model contributions, which makes this approach scalable to real-world situations of high-dimensional data.

4.3 Qualitative Analysis

A qualitative assessment of the suggested framework reveals considerable strengths that cannot be quantified in their entirety, particularly in terms of interpretability and soundness. The partitioning of feature space is a very central feature of clearing the heterogeneous attributes into individual categories according to the statistical attributes. This divide enables the domain expert to further analyse and comprehend the role played by the various subsets of features in the entire decision-making process of the model. The framework helps obtain a better understanding of the underlying data structure by grouping features with similar distributions or correlations, thereby allowing for more transparent analysis than traditional models, where each feature is considered on equal terms. This type of interpretability is particularly relevant in high-stakes areas, such as genomics or remote sensing, where the role of groups of features may inform additional research or interventions in the domain of interest.

Additionally, the dynamic ensemble aggregation system is highly robust due to its adaptive weighting mechanism. The framework is synchronized to keep track of changes or adjustments in the distributions of data by constantly calibrating the weight to each of the models relative to their levels of validation. This feature is essential in practical usage,

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 6, Issue 2, pp 94-102, February-2022

https://doi.org/10.47001/IRJIET/2022.602016

where data properties can change over time due to sensor conditions, biological variation, environmental samples, etc. The adaptive weights aid in the maintenance of reliable characteristics of prediction by prioritizing models that perform well in the prevailing circumstances and de-emphasizing models that do not. This concept helps alleviate the danger of going into negative performance as a result of concept drift. Furthermore, the need to constantly retrain is minimized using this dynamic adaptation, both saving on the amount of computational resources and preserving accuracy. Altogether, the qualitative advantages of the framework, such as better interpretability due to feature partitioning and resilience due to adaptive weighting, can be discussed as a complement to the framework's good quantitative performance, serving as an effective and practical solution to modelling complex, high-dimensional data.

V. CONCLUSION AND FUTURE WORK

The proposed multi-level ensemble framework, as outlined in this paper, is used to solve the problem of high-dimensional, heterogeneous feature distribution datasets that arise from these unique differences. The curse of dimensionality is a common issue when using traditional deep learning methods, in addition to the inability to account for the dissimilar statistical nature of different feature groupings. Our framework addresses these limitations in three major aspects, namely, feature space partitioning, adaptive deep models, and dynamic ensemble aggregation. The feature space partitioning step breaks the high-dimensional input into a collection of coherent subsets using statistical tests like divergence (Kullback-Leibler divergence) and correlation clustering to perform more specific and efficient modelling. Next, the auto-modelling approach, based on deep learning and adaptive models of each partition, learns specific feature representations using autoencoders, convolutional, and recurrent layers, allowing for dynamic adjustments to the data's heterogeneity. Lastly, the dynamic ensemble aggregation step combines the outputs of these different models in a weighted voting system, where the weights are adjusted according to validation performance. Such a joint mode substantially increases predictive accuracy, as illustrated in experiments on benchmark gene expression and hyperspectral imaging data, with reasonable computational costs compared to typical models and traditional ensemble techniques. Moreover, qualitative analysis sheds light on how feature partitioning enhances the interpretability of the present framework and how the weighting system is more robust, together making the framework quite useful in a real-life setting of working with high-dimensional data.

As this framework is extended and refined going forward, there are several promising directions to explore. Among the priorities are the adaptation of the model to real-time streaming data, where feature distributions can vary dynamically and new information appears constantly. The invention of online mechanisms for feature division and incremental model update would allow the framework to continue operating highly effectively in changing states, which is fundamental in most applications, such as sensor networks and financial prediction. Further, contemporary partitioning is based on supervised/semi-supervised statistical models; unsupervised feature partitioning methods should give the framework a level of flexibility not attained when labeled data is scarce. The introduction of new, complex clustering algorithms or unsupervised representation learning would be a welcome addition, allowing for the automatic identification of intrinsic feature groupings. Furthermore, adding explainability tools that will give more detailed specifics on model decisions at the partition and ensemble levels may enhance transparency and trustworthiness. Altogether, this multi-level ensemble paradigm sets a solid basis for further work towards viable, interpretable, and scalable deep learning solutions that fit the heterogeneity of high-dimensional data.

REFERENCES

[1] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. science, 313(5786), 504-507.

[2] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of machine learning research, 3(Mar), 1157-1182.

[3] Jolliffe, I. (2011). Principal component analysis. In International encyclopedia of statistical science (pp. 1094-1096). Springer, Berlin, Heidelberg.

[4] Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. Neural computation, 3(1), 79-87.

[5] Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10), 1345-1359.

[6] Dietterich, T. G. (2000, June). Ensemble methods in machine learning. In International workshop on multiple classifier systems (pp. 1-15). Berlin, Heidelberg: Springer Berlin Heidelberg.

[7] Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence, 35(8), 1798-1828.

[8] Chen, T., &Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 785-794).

[9] Kouw, W. M., & Loog, M. (2019). A review of domain adaptation without target labels. IEEE transactions on

ISSN (online): 2581-3048
Volume 6, Issue 2, pp 94-102, February-2022
https://doi.org/10.47001/IRJIET/2022.602016

pattern analysis and machine intelligence, 43(3), 766-785.

[10] Breiman, L. (1996). Bagging predictors. Machine learning, 24(2), 123-140.

[11] Freund, Y., &Schapire, R. E. (1997). A decision-theoretic generalization of online learning and an application to boosting. Journal of computer and system sciences, 55(1), 119-139.

[12] Caruana, R. (1997). Multitask learning. Machine learning, 28(1), 41-75.

[13] Zhang, Y., & Yang, Q. (2021). A survey on multi-task learning. IEEE transactions on knowledge and data engineering, 34(12), 5586-5609.

[14] Zheng, C., Wang, C., & Jia, N. (2019). An ensemble model for multi-level speech emotion recognition. Applied Sciences, 10(1), 205.

[15] Johnstone, I. M., & Titterington, D. M. (2009). Statistical challenges of high-dimensional data. Philosophical transactions of the Royal Society A: Mathematical, physical and engineering sciences, 367(1906), 4237-4253.

[16] Liu, R., & Gillies, D. F. (2016). Overfitting in linear feature extraction for classification of high-dimensional image data. Pattern Recognition, 53, 73-86.

[17] Nakamura-Zimmerer, T., Gong, Q., & Kang, W. (2021). Adaptive deep learning for high-dimensional Hamilton--Jacobi--Bellman equations. SIAM Journal on Scientific Computing, 43(2), A1221-A1247.

[18] Li, M., & Wang, Z. (2020). Deep learning for high-dimensional reliability analysis. Mechanical Systems and Signal Processing, 139, 106399.

[19] Gao, L., Song, J., Liu, X., Shao, J., Liu, J., & Shao, J. (2017). Learning in high-dimensional multimedia data: the state of the art. Multimedia Systems, 23(3), 303-313.

[20] Fang, S. H., & Wang, C. H. (2015). A novel fused positioning feature for handling the heterogeneous hardware problem. IEEE Transactions on Communications, 63(7), 2713-2723.

[21] Seni, G., & Elder, J. (2010). Ensemble methods in data mining: improving accuracy through combining predictions. Morgan & Claypool Publishers.

Citation of this Article:

Lalmohan Behera, & Venkataram Poosapati, “A Multi-Level Ensemble Framework for Adaptive Deep Learning in High-Dimensional Data Environments with Heterogeneous Feature Distributions” Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 6, Issue 2, pp 94-102, February 2022. Article DOI https://doi.org/10.47001/IRJIET/2022.602016

****