IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

H1B Visa Analysis using GANs

Mahi Maanas Reddy[1], Vaibhav Thalanki[2], Nitharshan CV[3], Shri Vaibbav R[4]
UG Students, School of Computer Science and Engineering[1,2,3,4]
Vellore Institute of Technology, Chennai, India

Abstract: *The H1B visa is a non-immigrant visa that enables foreign employees to enter the country and work there for a set amount of time. The purpose of the study is to analyze the 5 years of data on H1B Visa petitions from 2011-2016 and record the findings. Machine Learning algorithms will be employed to predict if the H1B Application Case status. The project also applies the functionality of Generative Adversarial Networks (GANs) to augment the training data. Tabular GANs are applied to the H1B Visa data and after augmentation, to test it on regular GANs, it is converted to images. These are fed to the generator network of a regular GAN model tested on MNIST data. The results show that GANs decrease the accuracy and increase the randomness of the data: Logistic Regression (before GAN=87.18%, after=73.8%), Random Forest (before=80.5%, after=70%), Gradient Boosting Ensemble (before=87.87%, after=71.3%), KNN (before=85.8%, after =71.4%), ANN (Error score Before GAN = 0.656, After GAN=1.033). Images formed by GANs to match the MNIST data showed satisfactory results after training for 5000+ epochs.*

Keywords: H1B visa.

I. INTRODUCTION

H1B visas are a type of non-immigrant, employment-based visa for temporary foreign workers in the US. A US firm must make a job offer to a foreign national before they may apply for an H1B visa and file a petition for one to the US immigration office[7]. As they finish their studies at a college or university and start working full-time, international students often apply for and maintain this visa status. The H1B visa is a non-immigrant visa that enables foreign employees to enter the country and work there for a set amount of time. Employing skilled foreign employees for specialized jobs like software developers, engineers, scientists, and healthcare experts is the most popular visa category employed by US businesses. The USCIS which stands for the U.S. Citizenship and Immigration Services is responsible for H1B visa program, which contains standards and restrictions that both the company and the individual must adhere to in order to be eligible[8]. The H1B visa program has generated discussion and controversy in recent years, with some claiming it displaces American residents from their employment and others claiming it is vital to the country's economy and creativity.

A deep learning model called a generative adversarial network (GAN) combines a discriminator and a generator neural network to produce artificial data that closely mimics the actual data. Although GANs are frequently used to create images and videos, their usage in creating tabular data has recently drawn a lot of interest. Structured data that is arranged in rows and columns, such as that found in databases or spreadsheets, is referred to as tabular data [1]. To create fresh, synthetic data for use in a variety of applications, including data augmentation, fraud detection, and predictive modeling, GANs may be trained on tabular data.

A significant advancement in machine learning is the use of GANs in tabular data applications since it enables the production of massive amounts of high-quality synthetic data, which can assist in overcoming issues related to data scarcity or privacy problems[3]. Nevertheless, there are particular difficulties with using GANs with tabular data, such as the requirement for specific loss functions and data preparation methods.

This project applies Tabular GANs on the H1B Visa Application dataset for augmentation. The data is doubled and then concatenated with the original data to produce a larger dataset. This is fed into ML (Machine Learning) algorithms like Logistic-Regression, Random-Forest, Gradient Boosting Ensemble, and K Nearest Neighbors. Data was also fed into a simple Artificial Neural Network (ANN) to analyze the error score [14]. Machine Learning algorithms will be

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

employed to predict if the H1B Application Case status is one of the following: Certified/Certified-withdrawn/Denied/Withdrawn/Pending-quality-review/Rejected/Invalidated [15].

II. LITERATURE SURVEY

[2] The paper by Qiao et al. (2020)uses transparent latent space controlled GANs - generative adversarial networks to provide a technique for renewable scenario generation. Authors' goal is to address the difficulties associated with integrating renewable energy sources into power systems, which calls for realistic and varied renewable energy scenarios. The suggested approach makes use of a configurable GAN to enable the production of scenarios depending on configurable inputs, such as the time of day and the weather. Additionally, the authors provide a visible latent space that enables the analysis of the situations created. The suggested method is compared to conventional scenario creation techniques and assessed on a real-world power system. Overall, the research makes a significant addition to the field of controlled GANs with transparent latent space for the development of sustainable energy scenario.

[4] The article addresses the potential applications of synthetic data, which is data that has been created artificially to address issues brought either by a lack of data or data of poor quality. The goal of the essay is to offer a survey which combines GANs and artificial data creation and serves as a good place for beginning researchers in the field. By querying four significant databases, the authors performed a review on state-of-the-art and comprehensively examined GANs, their most prevalent training issues, their most significant innovations, with a focus on GAN designs for tabular data. The importance of synthetic data and its ability to add value and safeguard privacy are emphasized in the article's conclusion. The article seeks to offer a broad overview of the most significant scientific advances as well as an organizational and chronological summary of the material covered in the publication.

[5]Using pairwise comparisons of data, the study offers a novel technique dubbed Pac GAN (Pairwise Comparison GAN), which enhances the training of Generative Adversarial Networks (GANs). Pac GAN allows for quicker and more stable convergence by using pairs of samples to direct the training of GANs. The authors demonstrate that this method is especially useful for high-resolution picture production tasks, which can be difficult for classic GAN training. They further show that for even higher performance, Pac GAN may be paired with other GAN methods, such as progressive growth. According to the results, Pac GAN beats a number of cutting-edge GAN training techniques on a number of datasets, including ImageNet and CelebA.The authors speculate that Pac GAN may find value in the creation of images and videos, as well as in other fields where pairwise comparisons might offer helpful training supervision signals.

[6] The paper begins by providing an introduction to the H1B visa program and its significance for the USA, followed by a literature survey on the relevant topics.The proposed approach involves the use of a decision tree algorithm for predicting the probability of an H1B visa application being approved. The data was collected from US Department of Labor's Office of Foreign Labor Certification website and pre-processed the data by removing duplicates, missing values, and outliers. They also performed feature selection using the chi-square test and mutual information gain. Evaluation metrics such as recall, accuracy, F1 score, recallwere calculated. The paper’s methodology performs better than existing methods. Overall, the paper provides a novel and effective approach for the prediction of H1B work visas in the USA using ML (machine learning). The authors have addressed an important issue and provided a solution that can potentially benefit both the US economy and the applicants seeking H1B visas.

[9] The paper reviews previous studies on work visas in the USA and the challenges of analyzing visa data, including data quality and legal complexities. The authors apply ML (machine learning) techniques, like decision-tree and logistic-regression, to analyze visa data and identify the factors that influence visa issuance. They find that factors such as the country of origin, industry, occupation, education level, and work experience significantly impact visa issuance, and decision tree models provide the most accurate predictions. The paper concludes by discussing the implications of their findings for policymakers and visa applicants, providing suggestions on how machine learning can improve visa processing systems, and highlighting the potential of machine learning techniques to enhance visa application outcomes. Overall, the paper contributes to the literature on work visas in the USA by leveraging ML techniques to analyze and understand visa issuance factors.

[10]The authors perform the comparison between different ML algorithms, such as decision-tree , random-forests, SVM, and logistic regression. The results show that the random forest algorithm outperforms the other models,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

achieving an accuracy of over 90% in predicting the outcome of H1B visa applications. The authors also provide an analysis of the important features that contribute to the prediction model, which includes factors such as occupation, employer location, and salary. In conclusion, the paper provides a useful way for predicting the H1B visa applications using ML algorithms. The results demonstrate the potential of these models to increase the accuracy and efficiency of the visa application process.

[11] The authors argue that existing GAN training methods suffer from issues such as instability, low quality, and lack of variation in generated images. To address these issues, the authors propose a technique that involves starting with a low-resolution image and gradually increasing its resolution during the training process. This allows the generator to learn and generate more complex and realistic images as the resolution increases. The paper presents experimental results demonstrating that the progressive growing technique leads to better quality, stability, and variation in generated images compared to existing methods. The authors also provide insights into the underlying reasons why this approach works well. Overall, the paper makes a significant contribution to the field of GANs by introducing a new training technique that overcomes some of the limitations of existing methods.

[12]A technique for assessing the effectiveness of handwritten Urdu text recognition systems by incorporating the learning experience of the MNIST dataset is suggested in the research paper. On a dataset of manually penned Urdu numbers and characters, they test their method, demonstrating that it performs more accurately than competing techniques. The article emphasizes how well transfer learning from the MNIST dataset works for other tasks that are similar to it and indicates that it can be used for other scripts and languages.

[13]This paper presents a Hierarchical Covering Algorithm (HCA). The CA can manage multi-category classification and large-scale data by building neural networks based on the traits of samples. The authors develop a hierarchical architecture based on fuzzy quotient space theory by applying the CA to create hidden nodes at the lowest level and defining a fuzzy equivalence relation R on upper spaces. The experiments conducted on the MNIST dataset show that the results show that the covering tree reflects the deep architecture of the problem and that the effects of a deep structure are superior to having a single level.

III. MATERIALS AND METHODS

3.1 Models Used

A. Artificial Neural Network

Due to their capacity to recognize intricate patterns in data and produce precise predictions, neural networks are frequently utilized for classification tasks. One input layer, two hidden layers, and a single output layer make up the components of the neural network. Each layer has nodes, commonly referred to as neurons, which are linked and use mathematical operations to process information.

The input layer of the neural network gets the data that has to be categorized. The output layer generates the classification result while the hidden layers analyze the input and extract pertinent characteristics.

The network is given a collection of labelled data during the training phase, and the biases, weights of the neurons are modified by backpropagation to reduce the discrepancy between the expected and actual output. This process keeps on until the network can correctly categorize brand-new, unheard-of data.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 50)                250

 dense_1 (Dense)             (None, 25)                1275

 dense_2 (Dense)             (None, 12)                312

 dense_3 (Dense)             (None, 1)                 13

=================================================================
Total params: 1,850
Trainable params: 1,850
Non-trainable params: 0
_________________________________________________________________
```

Fig 1: Model Summary of ANN

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

For a variety of classification tasks, including sentiment analysis, speech recognition, and picture classification, neural networks can be employed. They have shown to be quite successful in obtaining cutting-edge performance on a variety of difficult classification issues. The model summary is shown above:
The model has 3 layers and an output layer. Each layer is a densely connected network. The total trainable parameters for the network are 1,850. The model can be better understood by the following plot:

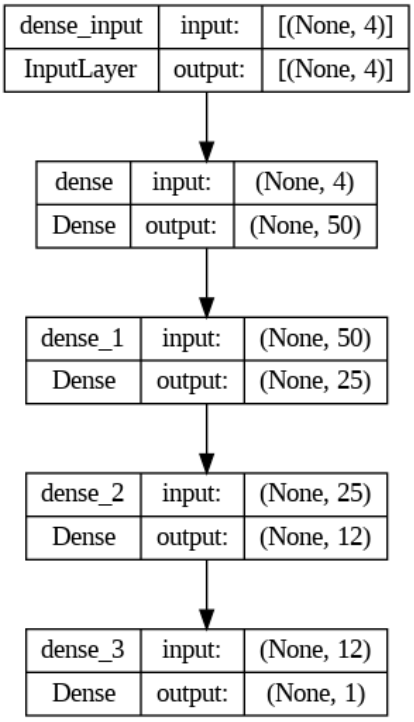

Fig 2: Model Summary Plot of ANN

This model is used to show the difference in the error score before and after applying Tab GAN (Tabular GANs) to our H1B data.

B. Logistic Regression

An analysis of the connection between dependent variable with one/many independent variables is done statistically using logistic regression. If the dependent variable is of type binary—that is, it may take one of two potential values, such as true or false, yes or no, or 0/1—it is employed in regression analysis.
A logistic function, that maps a real-valued number to a value between 0 and 1, is used in logistic regression to model the dependent variable as a function of the independent variables. Given a set of values for the independent variables, the logistic function is used to calculate the likelihood that the dependent variable equals 1.

Algorithm:

Parameters of hypothesis are initialized in random
Perform the logistic function to the linear hypothesis:

$$S(x) = \frac{1}{1 + e^{-x}}$$

Eq 1: Sigmoid function

Partial Derivative is calculated
Parameters are updated
Steps 2–4 are done again for n times (Stop when the cost converges)
Infer

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

C. Random Forest

ML tasks that require predicting a target variable when given some input characteristics are performed using the random forest method. It functions by merging the forecasts of many decision trees, each of which was trained using a portion of the input characteristics and training data that was randomly chosen. The pseudocode of the random forest algorithm is as follows:

Choose a portion of the input characteristics at random.

Create a decision tree using a random subset of the training data and the chosen features.

For a specific number of trees, repeat steps 1 and 2 once more.

To get a final forecast, combine the projections from all the trees.

$$RFfi_i = \frac{\Sigma_{j \, \epsilon \, all \, trees} \, normfi_{ij}}{T}$$

Eq 2: Random Forest Ensemble

D. Gradient Boosting Ensemble

A machine learning method called gradient boosting is utilized for regression and classification applications. In an ensemble approach, several weak learners—typically decision trees—are combined to produce a strong learner. In gradient boosting, decision trees are iteratively added to the ensemble, each one trained on the residuals of the previous trees, such that the final model predicts the target variable by combining the predictions of all the individual trees.

Algorithm:

initialize the model with a constant value
iterate for the specified number of trees
calculate the residuals of the current predictions
train a singular decision-tree on residuals
calculate the prediction of the current tree
update the model by adding the prediction of the current tree
Return the prediction of the final tree

E. K Nearest Neighbors

The ML method K Nearest Neighbors (KNN) is used for classification and regression applications. It is predicated on the premise that items with comparable goal values or objects that are adjacent to one another in the feature space are more likely to belong to the same class. A new input instance's class or target value is predicted by the K nearest neighbors algorithm based on the majority class or average target value of its neighbors.

Decide a distance metric (Example Euclidean, Manhattan)
iterate over all the test instances
calculate distance between test and all other training instances
find the K nearest neighbors
predict the class or target value of the test instance based on the neighbors (majority rule)
save the prediction

3.2 Dataset

The dataset is publicly available and open-source on Kaggle. This dataset includes information on H-1B petitions filed during a five-year period, totaling over 3 million records. “Employer Name, Case status, job description, case status, prevailing salary, workplace coordinates, occupation code, and year filed” are the attributes in the dataset. For the sake of the project, only the first 10K rows (after random shuffle) are taken. This is because anything more than that overwhelms the TabGAN model. However, the data is augmented 1.1 times the original and added back to the original data frame resulting in over 22K records.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

The Attributes of the dataset include:
Case Status (Final prediction attribute) - It can take any one of these values: Certified/ Certified-withdrawn/Denied/Withdrawn/Pending-quality-review/Rejected/Invalidated. It represents the conclusion of the H1B Visa Application.
Employer Name
SOC_NAME: `SOC_CODE` is the code according to the US Department of State.
Job title
Full-Time Position: Yes (1) or No (0)
Prevailing Wage: Wage for the job at given conditions
Year
Worksite
Latitude and Longitude of the worksite (for visualization purpose)

CASE_STATUS	EMPLOYER_NAME	SOC_NAME	JOB_TITLE	FULL_TIME_POSITION	PREVAILING_WAGE	YEAR	WORKSITE
CERTIFIED-WITHDRAWN	UNIVERSITY OF MICHIGAN	Biochemists And Biophysicists	Postdoctoral Research Fellow	N	36067.0	2016.0	Ann Arbor, Michigan
CERTIFIED-WITHDRAWN	GOODMAN NETWORKS, INC.	Chief Executives	Chief Operating Officer	Y	242674.0	2016.0	Plano, Texas
CERTIFIED-WITHDRAWN	PORTS AMERICA GROUP, INC.	Chief Executives	Chief Process Officer	Y	193066.0	2016.0	Jersey City, New Jersey
CERTIFIED-WITHDRAWN	GATES CORPORATION, A WHOLLY-OWNED SUBSIDIARY O...	Chief Executives	Regional Presiden, Americas	Y	220314.0	2016.0	Denver, Colorado
WITHDRAWN	PEABODY INVESTMENTS CORP.	Chief Executives	President Mongolia And India	Y	157518.4	2016.0	St. Louis, Missouri

Fig 3: 2011-2016 H1B VISA Application Data

MNIST Data:
The 70,000 handwritten digits in the MNIST dataset are individually represented by a 28x28 pixel grayscale picture. In the field ofML, this dataset is frequently used as a benchmark for image classification tasks. 10,000 photographs are included for testing, and 60,000 images are used for training.
Yann LeCun, Corinna Cortes, and Christopher Burges created the MNIST dataset to test the efficacy of machine-learning techniques for image recognition. Several ML methods, including as neural networks, SVM, KNN, and decision trees, have been developed and tested on the dataset. It is publicly available dataset and can be accessed from various sources including Kaggle.
The H1B augmented data will be converted to noisy random images which will be fed into the generator of the GAN network to recreate MNIST data.

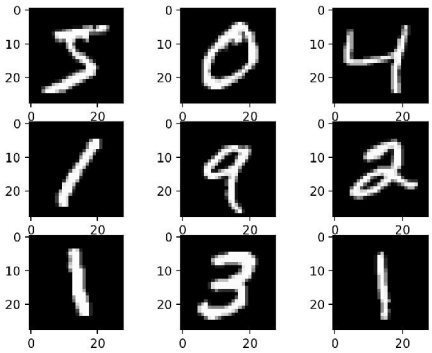

Fig 4: MNIST Data

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

3.3 Architecture

The overall Architecture for the project is shown below:

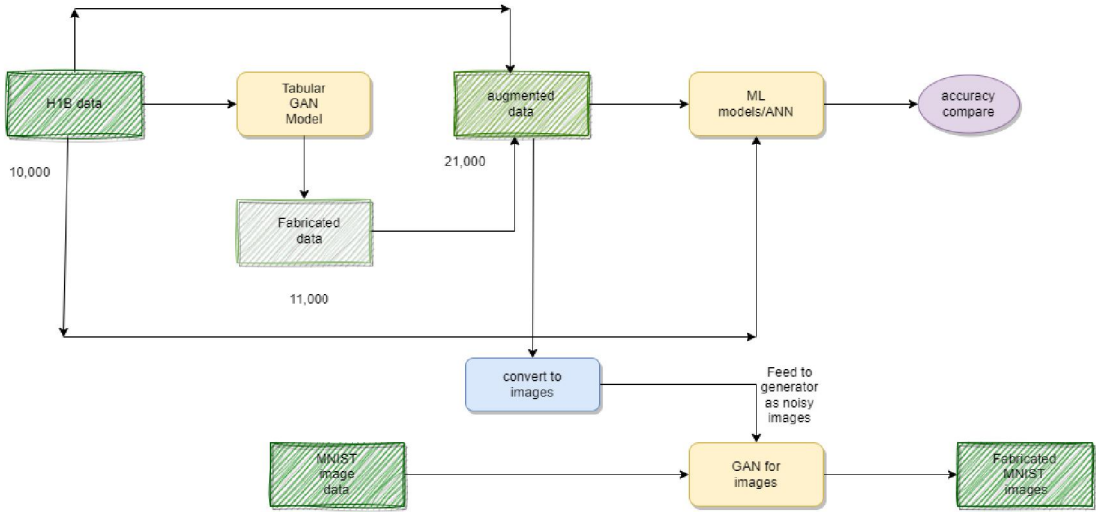

Fig 5: Overall Project Architecture

The GAN architecture is shown below:

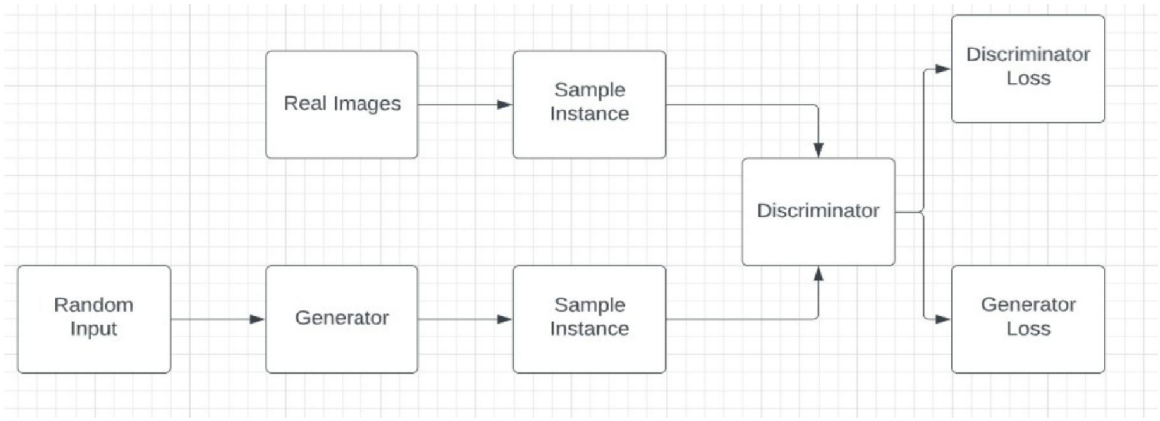

Fig 6: GAN Architecture

IV. NOVELTY

Tabular GANs

This project results in the application of Tabular GANs. The regular GAN is modified to handle tabular data. Usually, GANs augment and fabricate Image data like that of MNIST for example. Here, we use the functionality of GAN to augment H1B Visa Application Data which is tabular,

Another aspect of novelty in this project is to combine the Tab GAN with regular GANs. The regular GAN comprises of 2 networks namely, Generator and Discriminator. The Generator takes in random noise and tries to fool the discriminator network that it is a real image. The random noise here is replaced by the H1B augmented data (after Tab GAN). This is done by reshaping the numpy array and converting it into a greyscale image after normalizing every value in the array so that each value can act as a pixel.Notice that in Fig 6, instead of random input, we will be feeding the H1B augmented data after converting them into images. The MNIST image data will be fed into the block 'real images.

V. RESULTS AND DISCUSSION

5.1 Results

The accuracies of various Machine Learning models were recorded before and after the application of Tabular GANs on the H1B Visa Application dataset. The results show that GANs decrease the accuracy and increase the randomness

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

of the data: Logistic Regression (before GAN=87.18%, after=73.8%), Random Forest (before=80.5%, after=70%), Gradient Boosting Ensemble (before=87.87%, after=71.3%), KNN (before=85.8%, after =71.4%), ANN (Error score Before GAN = 1.06, After GAN=1.575).

Model Used	Accuracy / Error Score (*)	
	Before TabGAN	After TabGAN
ANN	1.060	1.575
Logistic Regression	87.18%	73.8%
Random Forest	80.5%	70%
Gradient Boosting ensemble	87.87%	71.3%
K Nearest Neighbors	85.8%	71.4%

Table 1: Performance metrics before and after GANs

Gradient Boosting Ensemble has the highest accuracy among all machine learning algorithms prior to TabGAN application. On the other hand, Logistic Regression has the highest after TabGAN application.

5.2 Image with explanations

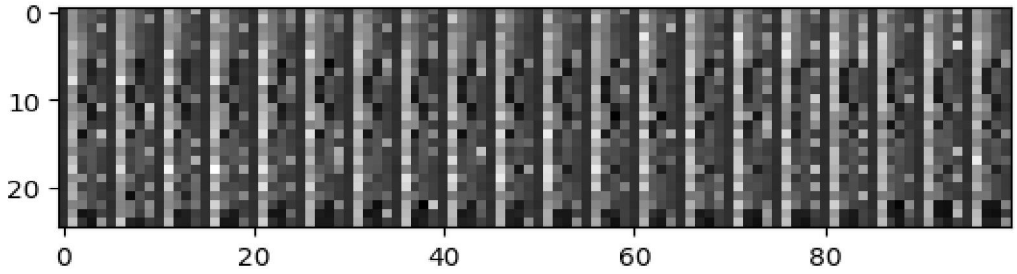

Fig 7: 25X100 TabGan Augmented H1B data

The augmented H1B data arising after TabGAN is converted into a numpy array with each sub array being 25x100 size. One of them is taken and represented as an image after normalization. This will serve as the input to the Generator in the GAN.

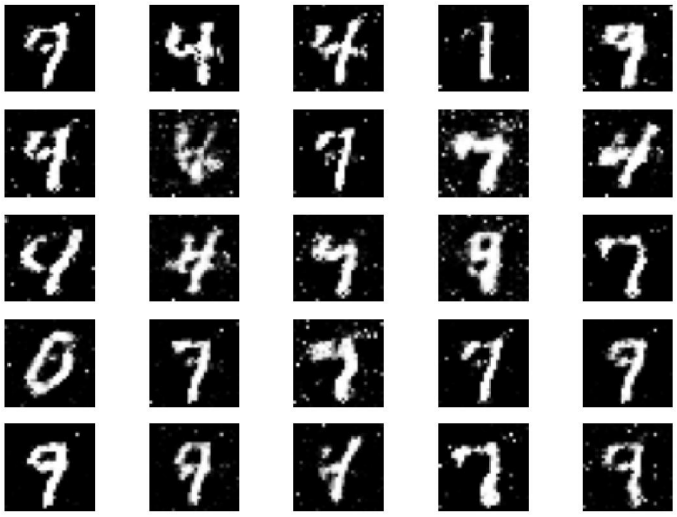

Fig 8: GAN result after 10K epochs

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

After training the GAN on the MNIST dataset for 10,000 epochs, the GAN then generated 25 images on its own replicating the MNIST database. As seen, each image can clearly be distinguished as a single digit by the human eye.

It is exceptional to see that our H1B data image (25x100 represented above) gets converted to each of these distinguishable digits after training GANs for 10,000 epochs. When the GAN was asked to spit out images after only 500 epochs, it looked something like this:

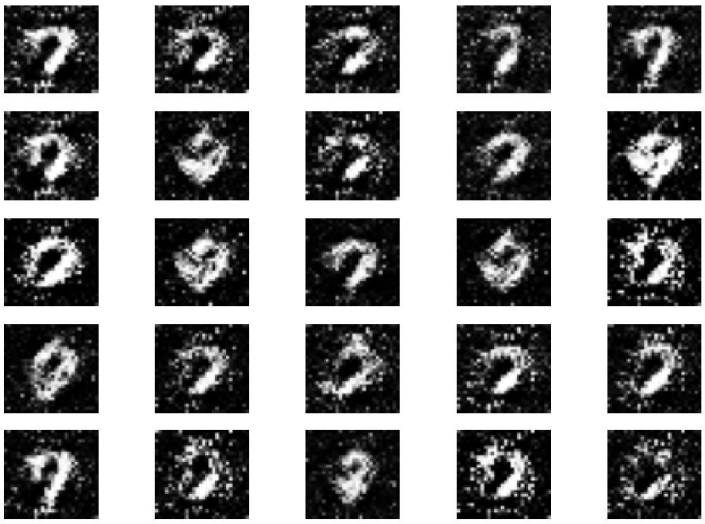

Fig 9: GAN Results after 500 epochs

VI. CONCLUSION

Tabular GANs were successfully applied to augment and fabricate H1B Visa Application data from 2011-2016. The data was preprocessed accordingly to be fed into the TabGAN. The augmented data consisted of 21,000 records compared to the original data of 10,000 records. Machine Learning Algorithms and ANNs were successfully tested on the data before and after manipulation using TabGANs. Gradient Boosting Ensemble has the highest accuracy among all machine learning algorithms prior to TabGAN application (87.87%). On the other hand, Logistic Regression has the highest after TabGAN application (73.8%). The MNIST data was successfully replicated by GANs after being fed normalized augmented 25x100 H1B Data as an image to act as the random image. After being run for 5000+ epochs, the results were satisfactory.

REFERENCES

[1] Wang, W., Wang, C., Cui, T., & Li, Y. (2020). Study of restrained network structures for wasserstein generative adversarial networks (WGANs) on numeric data augmentation. *IEEE Access*, *8*, 89812-89821.

[2] Qiao, J., Pu, T., & Wang, X. (2020). Renewable scenario generation using controllable generative adversarial networks with transparent latent space. *CSEE Journal of Power and Energy Systems*, *7*(1), 66-77.

[3] Khan, W., Zaki, N., Ahmad, A., Masud, M. M., Ali, L., Ali, N., & Ahmed, L. A. (2022). Mixed Data Imputation Using Generative Adversarial Networks. *IEEE Access*, *10*, 124475-124490.

[4]Figueira, A., & Vaz, B. (2022). Survey on synthetic data generation, evaluation methods and GANs. *Mathematics*, *10*(15), 2733.

[5] Lin, Z., Khetan, A., Fanti, G., & Oh, S. (2018). Pacgan: The power of two samples in generative adversarial networks. *Advances in neural information processing systems*, *31*.

[6]Thakur, Pooja & Singh, Mandeep & Singh, Harpreet & Rana, Prashant. (2018). An allotment of H1B work visa in USA using machine learning. International Journal of Engineering and Technology(UAE). 7. 93-103. 10.14419/ijet.v7i2.27.12642.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, August 2023

[7]Khaterpal*, R., Ahuja, H., Goel, J., Singh, K., … Manoj, R. (2020, May 30). Predicting the outcome of H-1B visa using ANN algorithm. International Journal of Recent Technology and Engineering (IJRTE). Blue Eyes Intelligence Engineering and Sciences Engineering and Sciences Publication - BEIESP. https://doi.org/10.35940/ijrte.a2917.059120

[8]Dombe, A., Rewale, R., & Swain, D. (2020). A Deep Learning-Based Approach for Predicting the Outcome of H-1B Visa Application. In Machine Learning and Information Processing (pp. 193-202). Springer, Singapore.

[9] Sundararaman, D., Pal, N., & Misraa, A. K. (2017). An analysis of nonimmigrant work visas in the USA using Machine Learning. International Journal of Computer Science and Security (IJCSS)

[10] D. Swain, K. Chakraborty, A. Dombe, A. Ashture and N. Valakunde, "Prediction of H1B Visa Using Machine Learning Algorithms," 2018 International Conference on Advanced Computation and Telecommunication (ICACAT), Bhopal, India, 2018, pp. 1-7, doi: 10.1109/ICACAT.2018.8933628.

[11] Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*.

[12] Ahmed, S. B., Hameed, I. A., Naz, S., Razzak, M. I., & Yusof, R. (2019). Evaluation of handwritten Urdu text by integration of MNIST dataset learning experience. *IEEE access*, *7*, 153566-153578.

[13] Chen, J., Zhao, S., & Zhang, Y. (2014). Hierarchical covering algorithm. *Tsinghua Science and Technology*, *19*(1), 76-81.

[14] Wei, L., Guan, L., & Qu, L. (2019). Prediction of sea surface temperature in the South China Sea by artificial neural networks. *IEEE geoscience and remote sensing letters*, *17*(4), 558-562.

[15] Khan, W., Zaki, N., Ahmad, A., Masud, M. M., Ali, L., Ali, N., & Ahmed, L. A. (2022). Mixed Data Imputation Using Generative Adversarial Networks. *IEEE Access*, *10*, 124475-124490.