EFFICIENT FASHION PREDICTION USING MNIST DATASET BY IMAGE CLASSIFICATION USING SUPPORT VECTOR MACHINE COMPARE WITH LOGISTIC REGRESSION WITH IMPROVED ACCURACY

S Gopi Ramesh[1], G. Charlyn Pushpa Latha[2*]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: Efficient novel fashion Prediction using MNIST dataset by image classification using a support vector machine and Logistic Regression.
Materials and Methods: This study contains 2 groups i.e Support Vector Machine and Logistic Regression. Each group consists of a sample size of 10 using G-power setting parameters: (α=0.05 and power=0.86) power value 0.4 respectively
Results: The Support Vector machine (SVM) is 91.2% which is more accurate than Logistic Regression (LR) of 72.6% in Fashion Forecasting and attained the significant value 0.094
Conclusion: The Support vector machine model is significantly better than the Logistic Regression in novel fashion Forecasting.

Keywords : Support Vector Machine, Logistic Regression, Accuracy, Novel Fashion Prediction, Deep Learning, Fashion Forecasting.

[1]Research Scholar, Department of Information Technology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu. India. Pincode: 602105.
[2*]Department of Information Technology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu. India, Pincode: 602105.

1. Introduction

MNIST's popularity stems from its small size, which allows deep learning researchers to easily test and prototype their methods. Additionally, all deep learning libraries and deep learning frameworks (e.g. Tensorflow, Pytorch) contain auxiliary functions (Albert-Weiss and Osman 2022). There are numerous functions and examples included with MNIST that you may utilize right away. A CNN-based LeNet-5 framework is recommended for training CNN variables on the Fashion MNIST dataset (Zaidi et al. 2021). The findings of the experiments demonstrate that LeNet-5 is effective. The accuracy of the model was over 98%. As a result, it outperforms both the original CNN model and additional state-of-the-art models in the literature. (Hara, Jagadeesh, and Piramuthu 2016) The aim of recognition is clothing worn in natural situations, the challenge is extended to include garment detection. To facilitate classification, solutions for these types of challenges often aim to identify the relevant locations for each piece of clothing. (Rocamora and O'Neill 2008). Application of Fashion trend forecasting is a multi-industry technique that includes autos, pharmaceuticals, food and drinks, literature, and home furnishings. Fashion Forecasting are in charge of enticing customers and assisting retailers and designers in selling their products (Song, Nie, and Wang 2019).

There are about 125 articles in IEEE xplore and in 130 Scopus related to this study. In a study by (Blaszczyk and Wubs 2018). This work will look at how novel fashion prediction communication is evolving, as well as its function in generating fashion ideas and pushing trends. The focus is on the generation that has never known life without the internet, with particular attention paid to the online content they consume and the impact it has on their fashion choices. (Amritha and Suresh 2020) explains that Business executives are on the front foot, attempting to innovate while continuing to engage their core constituencies, despite the fashion industry's record-low economic profitability. Many organizations are reconnecting with their supply chains as a result of the recent interruptions. (Parsons 2018) implements different novel Fashion prediction looks for ready to wear during winter as a winter collection.

Our institution is keen on working on latest research trends and has extensive knowledge and research experience which resulted in quality publications (Rinesh et al. 2022; Sundararaman et al. 2022; Mohanavel et al. 2022; Ram et al. 2022; Dinesh Kumar et al. 2022; Vijayalakshmi et al. 2022; Sudhan et al. 2022; Kumar et al. 2022; Sathish et al. 2022; Mahesh et al. 2022; Yaashikaa et al. 2022). Although detecting a visual object from a photograph is a straightforward task for a human, it is tremendously difficult for a computer system to do it with human-level accuracy (Leite Arnø, Godhavn, and Aamo 2021). The approach must be resistant to a wide range of changes in order to appropriately recognize and categorize the photos (Yang et al. 2022). Various lighting circumstances, size and perspective changes, deformations, and occlusions, for example, could cause the algorithm to predict the picture class erroneously. The purpose of parametric regression analysis is to come up with a way to identify Fashion MNIST photos using deep learning convolutional neural network variants. Parametric regression analysis, which is referred to simply as regression analysis in this section, is the most common type of regression analysis (Ertl et al. 2021).

2. Materials and Methods

The proposed study was conducted in the DBMS Laboratory of the Saveetha School of Engineering under the supervision of faculty. To conduct this study, two groups were formed. Table 1 shows that group 1 is SVM and group 2 is Linear regression. The sample size was generated using clinical analysis, with G power set to 80%, 10 sample sizes estimated for each group, totaling 20, 94 percent confidence, pretest power set to 80%, and enrollment ratio set to The Support Vector machine (SVM) is 91.2% which is more accurate than Logistic Regression (LR) of 72.6% The accuracy of two classifiers, SVM and Linear regression, was compared. Image, vocabulary, color, and image size variables are independent factors in images. Images and clothes are dependent variables. Outfits is an Independent variable.

Support Vector Machine

Support Vector Machine (SVM) is utilized as a classifier. SVM is a regulated AI calculation which characterizes the capacity that orders information into two classes. In the proposed framework, have characterized two classes as dangerous or non-malignant. SVM is a twofold order strategy that takes as information from two classes and yields a model document for arranging obscure or known information into one of two classes. Preparing a SVM includes taking care of known information to the SVM alongside recently released choice qualities, in this way shaping a preparation Set. It is from the preparation set that a SVM gets its knowledge to group obscure information.

Pseudocode

Input: Fashion prediction_Input Features
To forecast fashion, assign training and testing datasets.
Output: Fashion predictions are classified.
Function: Support_Vector_Machine(Input features F, Label vector V=[1.....n])
Step 1: Choose the ideal cost and gamma factor.
Step 2: do while (conditioning)
Step 3: Execute the training step for every predetermined set of input file features.
Step 4: Execute the classification process for a predetermined number of input file features.
Step 5: End the process while
Step 6: Classify the prediction performance.

Logistic Regression
The binary classification method is called Logistic Regression (LR). The relationship between the anticipated variable and additionally at least one independent component is quantified by logistic regression (our features). The sigmoid function is then used to convert the obtained probabilities to binary values. The Sigmoid-Function is a function that standardizes numbers between the range of 0 and 1, but never exactly at those bounds, given a real valued input. We then re-transform the data between 0 and 1 to either 0 or 1 by using a transformer predictor.

Pseudocode For Linear Regression
Input: Fashion prediction _Input Features
Assign Training and Testing Dataset of fashion prediction
Output: Classification on fashion prediction
Function: Logistic Regression (Input features I)
Step 1: Consider matrix M for Input features I
Step 2: Configure the Regressor parameter θ
Step 3: Add the Target Class Vector C= [1... N]
Step 4: while (condition) do
Step 5: Generate a binary vector B
Step 6: If it belongs to the class label C, check each class label C where B=1.
Step 7: where B=0, if not belong to the class label C
Step 8: Apply Logistic Regression to the Matrix M in order to discover the parameter regressor θ
End while
Return prediction outcomes

Statistical Analysis
The statistical analysis was performed using the software program Statistical Package for the Social Sciences, Version 26. For accuracy, a separate sample T-test was performed. Using the SPSS Software tool, standard deviation and standard mean errors were also computed. Table 3 displays the significance scores for proposed and current algorithms. It contains the aggregate statistical values of suggested and implemented algorithms.

3. **Results**

The statistical analysis of the two groups Support Vector Machine (group 1) has a mean accuracy that is higher than Logistic Regression (LR) (group 2), and Support Vector Machine (SVM) has a standard error mean that is marginally lower, as shown in table 3. As indicated in Table 2, the accuracy scores for the Support Vector Machine algorithm were 91.2% and for Logistic Regression were 82.6%. Table 4 displays the outcomes of the Independent Sample Tests with a 95% confidence interval and a significance level of 0.05. Each algorithm's average accuracy is determined after the algorithms are tested using 6 different sample sizes. With standard deviation error shown for both algorithms, Fig. 1 shows a bar chart of accuracy.

4. **Discussion**

The study's significance value is 0.094, indicating that Generator utilizing Generative Adversarial looks to be superior to Support Vector Machine. Linear Regression surpasses the Support Vector Machine Algorithm in determining the appropriate innovative fashion Forecasting Type using the encoder-decoders model. The information is obtained through a series of iterations in order to find different scales of accuracy rates. A statistical analysis was utilized to do an independent sample t-test and analyze the accuracy outcomes based on the accuracy rates (Vu et al. 2022). The mean accuracy of a Support Vector Machine novel Fashion prediction is 91 percent, while a Linear Regression Algorithm forecast is 76 percent.

Similar researchers have looked at fashion forecasting with deep learning approaches. Those discovered, on the other hand, have concentrated on utilizing regression models to forecast sales figures. Complex deep learning methods were frequently used in these models (Shetty, Shales, and Narayan 2022). Other research aimed to improve the performance of the learning models that had been investigated. The opposite was addressed in the context of a scenario in which novel fashion prediction retailers were confronted with highly unpredictable demand for their items. Due to the unpredictability of demand, shops were left with either excess inventory or stock-outs, causing economic problems (Alig et al. 2022).

The disadvantage of following fashion Forecasting is that it may be harmful to our environment. In truth, every material thing that must be made pollutes the environment and frequently contributes to global warming. Excessive consumption also contributes to significant resource depletion (Parodis, Tamirou, and Houssiau 2022). The system's future aim is to improve, to cover a larger number of photos while taking less time in training the data set. As a result of characteristics like these, accuracy and exact precision values can be increased.

5. Conclusion

Based on the experimental results, Support Vector machine (SVM) has been proved to Predict novel fashion prediction more significantly than Logistic Regression (LR). As a result, we've created a multi-step framework that combines image segmentation, face recognition, and logistic regression into a single phase. We can extract picture attributes from a raw image using image segmentation and face recognition techniques. The Support Vector machine (SVM) is 91.2% which is more accurate than Logistic Regression (LR) of 72.6% . This picture recognition and segmentation technique allows us to concentrate on the image's most important aspects, allowing us to do further analysis. We may obtain information and data from these photos using these features, and the data can then be used for further learning analysis.

DECLARATIONS

Conflicts Of Interest

No conflicts of interest in this manuscript.

Author Contributions

Author GR was involved in data collection, data analysis, data extraction, manuscript writing.Author CPG was involved in conceptualization, data validation, and critical review of the manuscript.

Acknowledgement

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

1. Vee Eee Technologies Solution Pvt. Ltd., Chennai.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

6. References

Albert-Weiss, Dominique, and Ahmad Osman. 2022. “Interactive Deep Learning for Shelf Life Prediction of Muskmelons Based on an Active Learning Approach.” Sensors 22 (2). https://doi.org/10.3390/s22020414.

Alig, Annabel H. S., Volker Heinemann, Michael Geissler, Ludwig Fischer von Weikersthal, Thomas Decker, Kathrin Heinrich, Swantje Held, et al. 2022. “Exact Primary Tumor Location in mCRC: Prognostic Value and Predictive Impact on Anti-EGFR mAb Efficacy.” Cancers 14 (3). https://doi.org/10.3390/cancers14030526.

Amritha, B., and Kalyani Suresh. 2020. “Sustainability Is the New Black: Exploring Website Communication Practices of Indian Sustainable Fashion Brands.” Fashion, Style & Popular Culture. https://doi.org/10.1386/fspc_00042_1.

Blaszczyk, Regina Lee, and Ben Wubs. 2018. The Fashion Forecasters: A Hidden History of Color and Trend Prediction. Bloomsbury Publishing.

Dinesh Kumar, M., V. Godvin Sharmila, Gopalakrishnan Kumar, Jeong-Hoon Park, Siham Yousuf Al-Qaradawi, and J. Rajesh Banu. 2022. “Surfactant Induced Microwave Disintegration for Enhanced Biohydrogen Production from Macroalgae Biomass: Thermodynamics and Energetics.” Bioresource Technology 350 (April): 126904.

Ertl, Peter, Grégori Gerebtzoff, Richard Lewis, Hagen Muenkler, Nadine Schneider, Finton Sirockin, Nikolaus Stiefl, and Paolo Tosco. 2021. “Chemical Reactivity Prediction: Current Methods and Different Application Areas.” Molecular Informatics, December, e2100277.

Hara, Kota, Vignesh Jagadeesh, and Robinson Piramuthu. 2016. “Fashion Apparel Detection: The Role of Deep Convolutional Neural Network and Pose-Dependent Priors.” 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). https://doi.org/10.1109/wacv.2016.7477611.

Kumar, J. Aravind, J. Aravind Kumar, S. Sathish, T. Krithiga, T. R. Praveenkumar, S. Lokesh, D. Prabu, A. Annam Renita, P. Prakash, and M. Rajasimman. 2022. “A Comprehensive Review on Bio-Hydrogen Production from Brewery Industrial Wastewater and Its Treatment Methodologies.” Fuel. https://doi.org/10.1016/j.fuel.2022.123594.

Leite Arnø, Mikkel, John-Morten Godhavn, and Ole Morten Aamo. 2021. “A Divided and

Prioritized Experience Replay Approach for Streaming Regression.” MethodsX 8 (November): 101571.

Mahesh, Narayanan, Srinivasan Balakumar, Uthaman Danya, Shanmugasundaram Shyamalagowri, Palanisamy Suresh Babu, Jeyaseelan Aravind, Murugesan Kamaraj, and Muthusamy Govarthanan. 2022. “A Review on Mitigation of Emerging Contaminants in an Aqueous Environment Using Microbial Bio-Machines as Sustainable Tools: Progress and Limitations.” Journal of Water Process Engineering. https://doi.org/10.1016/j.jwpe.2022.102712.

Mohanavel, Vinayagam, K. Ravi Kumar, T. Sathish, Palanivel Velmurugan, Alagar Karthick, M. Ravichandran, Saleh Alfarraj, Hesham S. Almoallim, Shanmugam Sureshkumar, and J. Isaac JoshuaRamesh Lalvani. 2022. “Investigation on Inorganic Salts K2TiF6 and KBF4 to Develop Nanoparticles Based TiB2 Reinforcement Aluminium Composites.” Bioinorganic Chemistry and Applications 2022 (January): 8559402.

Parodis, Ioannis, Farah Tamirou, and Frédéric A. Houssiau. 2022. “Treat-to-Target in Lupus Nephritis. What Is the Role of the Repeat Kidney Biopsy?” Archivum Immunologiae et Therapiae Experimentalis 70 (1): 8.

Parsons, Jean L. 2018. “The Shirtwaist: Changing the Commerce of Fashion.” Fashion, Style & Popular Culture. https://doi.org/10.1386/fspc.5.1.7_1.

Ram, G. Dinesh, G. Dinesh Ram, S. Praveen Kumar, T. Yuvaraj, Thanikanti Sudhakar Babu, and Karthik Balasubramanian. 2022. “Simulation and Investigation of MEMS Bilayer Solar Energy Harvester for Smart Wireless Sensor Applications.” Sustainable Energy Technologies and Assessments. https://doi.org/10.1016/j.seta.2022.102102.

Rinesh, S., K. Maheswari, B. Arthi, P. Sherubha, A. Vijay, S. Sridhar, T. Rajendran, and Yosef Asrat Waji. 2022. “Investigations on Brain Tumor Classification Using Hybrid Machine Learning Algorithms.” Journal of Healthcare Engineering 2022 (February): 2761847.

Rocamora, Agnès, and Alistair O’Neill. 2008. “Fashioning the Street: Images of the Street in the Fashion Media.” Fashion as Photograph. https://doi.org/10.5040/9780755696420.ch-013.

Sathish, T., V. Mohanavel, M. Arunkumar, K. Rajan, Manzoore Elahi M. Soudagar, M. A. Mujtaba, Saleh H. Salmen, Sami Al Obaid, H. Fayaz, and S. Sivakumar. 2022. “Utilization of Azadirachta Indica Biodiesel, Ethanol and Diesel Blends for Diesel Engine Applications with Engine Emission Profile.” Fuel. https://doi.org/10.1016/j.fuel.2022.123798.

Shetty, Varun, Sufina Shales, and Pradeep Narayan. 2022. “Lactate Levels: A Reliable Tool to Predict Mortality in Acute Type A Aortic Dissections?” Journal of Cardiac Surgery, February. https://doi.org/10.1111/jocs.16322.

Song, Xuemeng, Liqiang Nie, and Yinglong Wang. 2019. Compatibility Modeling: Data and Knowledge Applications for Clothing Matching. Morgan & Claypool Publishers.

Sudhan, M. B., M. Sinthuja, S. Pravinth Raja, J. Amutharaj, G. Charlyn Pushpa Latha, S. Sheeba Rachel, T. Anitha, T. Rajendran, and Yosef Asrat Waji. 2022. “Segmentation and Classification of Glaucoma Using U-Net with Deep Learning Model.” Journal of Healthcare Engineering 2022 (February): 1601354.

Sundararaman, Sathish, J. Aravind Kumar, Prabu Deivasigamani, and Yuvarajan Devarajan. 2022. “Emerging Pharma Residue Contaminants: Occurrence, Monitoring, Risk and Fate Assessment – A Challenge to Water Resource Management.” Science of The Total Environment. https://doi.org/10.1016/j.scitotenv.2022.153897.

Vijayalakshmi, V. J., Prakash Arumugam, A. Ananthi Christy, and R. Brindha. 2022. “Simultaneous Allocation of EV Charging Stations and Renewable Energy Sources: An Elite RERNN-m2MPA Approach.” International Journal of Energy Research. https://doi.org/10.1002/er.7780.

Vu, Lan Phuong, Maria Zyulina, Alexandra Hingst, Gregor Schnakenburg, and Michael Gütschow. 2022. “Combinatorial Assembly, Traceless Generation and in Situ Evaluation of Inhibitors for Therapeutically Relevant Serine Proteases.” Bioorganic Chemistry 121 (February): 105676.

Yaashikaa, P. R., P. Senthil Kumar, S. Jeevanantham, and R. Saravanan. 2022. “A Review on Bioremediation Approach for Heavy Metal Detoxification and Accumulation in Plants.” Environmental Pollution 301 (May): 119035.

Yang, Ziduo, Weihe Zhong, Qiujie Lv, and Calvin Yu-Chian Chen. 2022. “Multitask Deep Learning with Dynamic Task Balancing for Quantum Mechanical Properties Prediction.” Physical Chemistry Chemical Physics: PCCP, February. https://doi.org/10.1039/d1cp05172e.

Zaidi, Syed Mohammed Arshad, Varun Chandola, Muhanned Ibrahim, Bianca Romanski, Lucy D. Mastrandrea, and Tarunraj Singh. 2021. "Multi-Step Ahead Predictive Model for Blood Glucose Concentrations of Type-1 Diabetic Patients." Scientific Reports 11 (1): 24332.

Tables and Figures

Table 1. Group, Accuracy, and Loss value uses 8 columns with 8 width data for novel image caption generators.

SI.NO	Name	Type	Width	Decimal	Columns	Measure	Role
1	Group	Numeric	8	2	8	Nominal	Input
2	Accuracy	Numeric	8	2	8	Scale	Input
3	Loss	Numeric	8	2	8	Scale	Input

Table 2. Accuracy and Loss Analysis of SVM and Logistic Regression.

S.No	GROUPS	ACCURACY	LOSS
1	SVM	91.00	9.00
		81.68	18.32
		74.56	25.44
		86.25	13.75
		78.64	21.36
		85.78	14.22
		68.94	31.06
		90.56	9.44
		84.36	15.64

		76.25	**23.75**
2	**Logistic Regression**	**72.00**	**22.00**
		67.21	**32.79**
		61.78	**38.22**
		73.56	**26.44**
		63.75	**36.25**
		59.14	**40.86**
		57.56	**42.44**
		75.12	**24.88**
		60.53	**39.47**
		56.85	**43.15**

Table 3. Group Statistic analysis, representing Support Vector Machine (mean accuracy 91.2%, standard deviation 0.76274) and Logistic Regression (mean accuracy 82.6%, standard deviation 0.44664).

	GROUP	**N**	**Mean**	**Std.Deviation**	**Std.Error Mean**
ACCURACY	**SVM**	**10**	**90.2200**	**0.76274**	**0.24120**
	Logistic Regression	**10**	**72.6520**	**0.44664**	**0.14124**
LOSS	**SVM**	**10**	**18.1970**	**7.12108**	**2.23211**
	Logistic Regression	**10**	**34.6320**	**6.71992**	**2.44635**

Table 4. Independent Sample Tests results with confidence interval as 95% and level of significance as 0.05 (Support Vector Machine appears to perform significantly better than Logistic Regression with the value of p=0.09).

		F	**Sig.**	**t**	**df**	**Mean Diffencen**	**Std.Error difference**	**Lower**	**Upper**

						e			
ACCURACY	**Equal variances assumed**	**3.118**	**0.094**	**30.654**	**18**	**8.56800**	**0.27951**	**7.980777**	**9.15523**
	Equal Variances not assumed	**_**	**_**	**30.654**	**14, 523**	**8.56800**	**0.27951**	**7.97053**	**9.15523**
LOSS	**Equal variances assumed**	**3.118**	**0.094**	**-30.654**	**18**	**-8.56800**	**3.33091**	**-23.44999**	**-9.45401**
	Equal Variances not assumed	**_**	**_**	**-30.654**	**14, 523**	**-8.56800**	**3.33091**	**-23.45276**	**-9.45124**

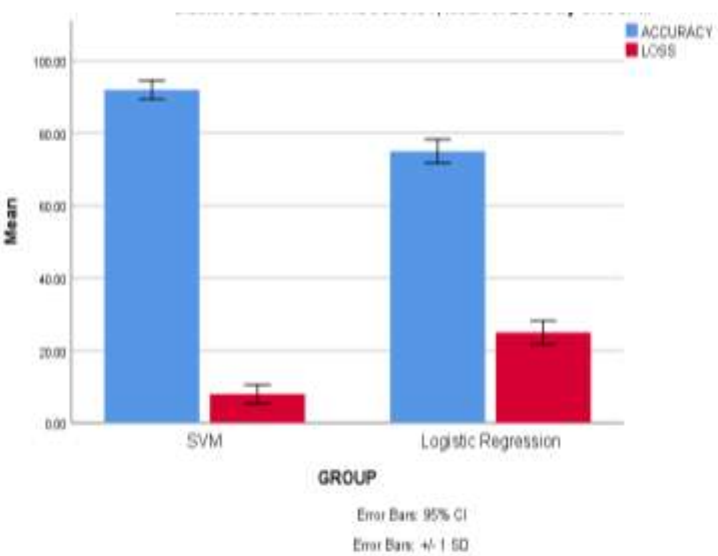

Fig. 1. Comparison of Support vector machine and Logistic Regression in terms of accuracy. The mean accuracy of the Support vector machine is greater than Logistic Regression and standard deviation is also slightly higher than Logistic Regression. X-axis: Support Vector Machine vs Logistic Regression. Y-axis: Mean accuracy of detection + 1 SD.