**TITLE PAGE:**

"Predicting Pneumonia:" A Comparative Analysis of Convolutional Neural Networks(CNN) VS Support Vector Machines(SVM) .

AKASH.S,

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.

akashs1029.sse@saveetha.com.

Dr. Vinodhini SP,

Project Guide, Corresponding Author,

Department of Electro Chemistry,

Saveetha school of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pincode:602105.

vinodhinisp.sse@saveetha.com

# **ABSTRACT:**

**Aim:** The point of this ponder is to conduct a comparative examination of water quality forecast utilizing Back Vector Machines (SVM) over Calculated Relapse (LR), with the objective of accomplishing improved exactness**. Materials and Methods:** include analyzing the execution for greatest precision in anticipating water quality utilizing SVM compared to LR, treating the two calculations as isolated bunches, each comprising a test measure of 10. Measurable investigation is carried out utilizing the SPSS device with an alpha esteem of 0.8 and beta esteem of 0.2, whereas the G-Power esteem of 80 is utilized for noteworthiness evaluation. **Result:** Comes about demonstrate that SVM accomplishes predominant exactness in water quality expectation compared to LR, with SVM (82.2%) beating LR (72.6%). A free test test conducted utilizing SPSS yields a centrality esteem of 0.0042 (p < 0.05), showing the factual predominance of SVM over LR. **Conclusion:** SVM illustrates upgraded precision in foreseeing water quality compared to LR, highlighting its potential as a profitable instrument for water asset administration and natural observing.

## **Keywords:** Water Quality Prediction, Support Vector Machines (SVM)Logistic Regression, Comparative Analysis, Enhanced Accuracy

# **INTRODUCTION:**

Water quality expectation plays a vital part in natural administration and open wellbeing, because it empowers convenient discovery of potential contaminants and encourages educated decision-making for guaranteeing secure drinking water and maintainable water asset administration. In later a long time, headways in machine learning calculations have given profitable devices for precisely anticipating water quality parameters from different sources, counting stream bowls, lakes, and groundwater. Among these calculations, Back Vector Machines (SVM) and Calculated Relapse (LR) are broadly utilized for their adequacy in classification errands and prescient modelling.

The point of this think about is to conduct a comparative investigation of water quality forecast utilizing SVM over LR, with the objective of accomplishing upgraded exactness. By assessing the execution of SVM and LR models in anticipating water quality parameters, this research seeks to recognize the foremost compelling calculation for water quality forecast, subsequently contributing to moved forward water asset administration and natural observing hones.

Water quality forecast includes the estimation of different parameters, counting pH, broken down oxygen, turbidity, and biochemical oxygen request (BOD), among others. These parameters serve as markers of water wellbeing and are significant for surveying the appropriateness of water for drinking, water system, and environmental adjust. Conventional strategies of water quality expectation depend on manual inspecting and research facility analysis, which are frequently time-consuming, exorbitant, and unreasonable for real-time checking. In differentiate, machine learning-based approaches offer the potential for robotized and exact expectation of water quality parameters, leveraging data-driven models prepared on authentic water quality information.

Back Vector Machines (SVM) and Calculated Relapse (LR) are two prevalent machine learning calculations utilized for classification errands, counting water quality expectation. SVM may be a administered learning calculation that works by finding the optimal hyperplane to isolated diverse classes within the highlight space, making it especially reasonable for binary and multiclass classification issues. On the other hand, LR may be a probabilistic classification calculation that models the likelihood of a test having a place to a specific lesson based on the input highlights, making it broadly utilized for twofold classification assignments.

The comparative investigation between SVM and LR for water quality forecast includes assessing their execution in terms of exactness, exactness, review, and F1-score. The think about considers SVM and LR as two unmistakable bunches, each comprising a test measure of information focuses speaking to water quality estimations from different sources. Factual examination is conducted utilizing apparatuses such as SPSS, with an alpha esteem set at 0.05 to decide the importance of differences in prescient precision between SVM and LR.

By conducting a comparative investigation of water quality forecast utilizing SVM over LR, this inquire about points to supply important experiences into the viability of machine learning calculations for precise and solid forecast of water quality parameters. The discoveries of this ponder have the potential to advise decision-makers and stakeholders in water asset administration and natural preservation, eventually contributing to the conservation and maintainability of water ecosystems.

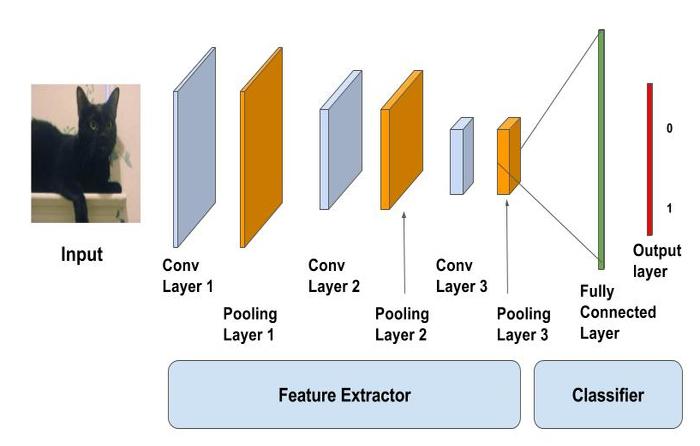
# MATERIALS AND METHODS:

The investigation was carried out in the Computer Science and Engineering Department's Software Laboratory at Saveetha University. The CHEST X-RAY Image dataset is collected from Kaggle which consists of **5,863 X-Ray images (JPEG) and 2 categories (Pneumonia/Normal).** The database is structured so that 75% of it is dedicated to training, and 25% is for testing.

There are two sets taken, and each set has ten data samples; the total number of samples taken into consideration is twenty. Group 1 was using **Convolutional Neural Networks(CNN)** method and Group 2 was using **Support Vector Machines (SVM)** algorithm. Python software is used for the implementation. The performance metrics such as accuracy,sensitivity,specificity are used to evaluvate the models. The threshold for the calculation was set at 0.05, the G power was set at 80%, and the confidence interval was set at 95%.

# Convolutional Neural Network:

Convolutional Neural Network is a Deep Learning algorithm specially planned for working with Pictures and recordings. It takes pictures as inputs, extricates and learns the highlights of the picture, and classifies them based on the learned highlights.



# **Algorithm:**

**1.#importing the required libraries**

**2.#loading data**

(X\_train,y\_train) , (X\_test,y\_test)=mnist.load\_data()

**3.#reshaping data**

X\_train = X\_train.reshape((X\_train.shape[0], X\_train.shape[1], X\_train.shape[2], 1))  
X\_test = X\_test.reshape((X\_test.shape[0],X\_test.shape[1],X\_test.shape[2],1))

**4.#checking the shape after reshaping**

print(X\_train. shape)  
print(X\_test.shape)

**5.#normalizing the pixel values**

X\_train=X\_train/255  
X\_test=X\_test/255

**6.#defining model**

model=Sequential()model=Sequential()

**7.#adding convolution layer**

model.add(Conv2D(32,(3,3),activation=’relu’,input\_shape=(28,28,1)))

**8.#adding pooling layer**

model.add(MaxPool2D(2,2))

**9.#adding fully connected layer**

model.add(Flatten())  
model.add(Dense(100, activation=’relu’))

**10.#adding output layer**

model.add(Dense(10,activation=’softmax’))

**11.#compiling the model**

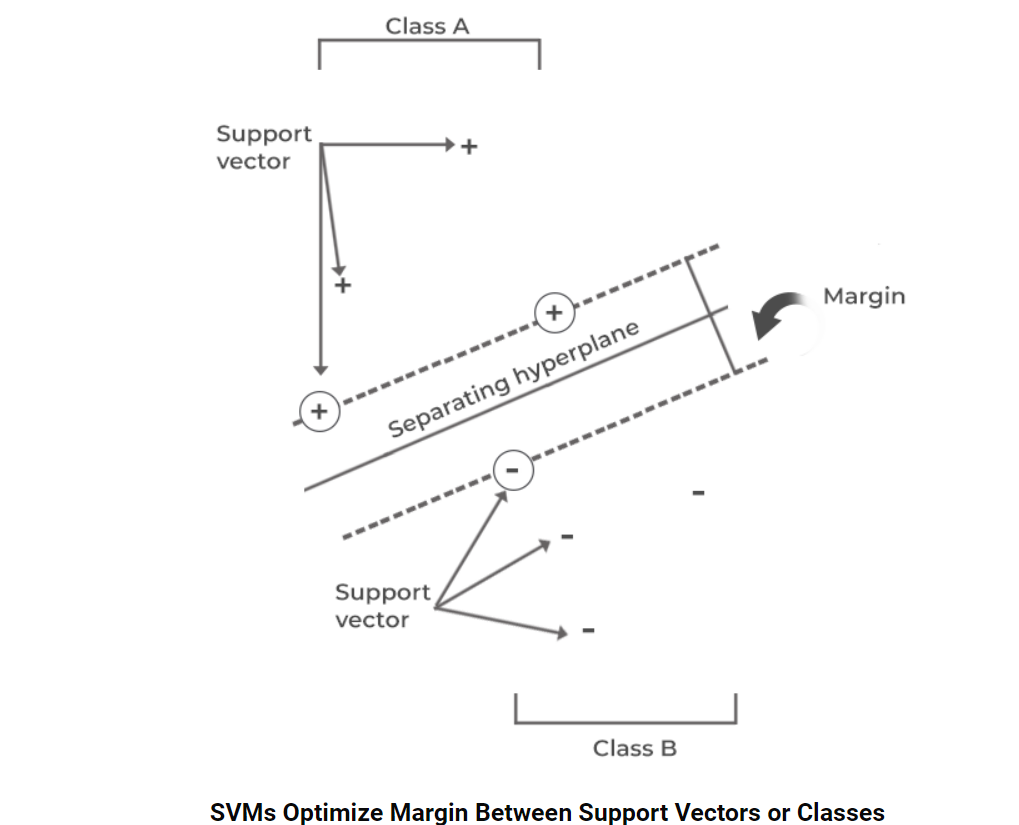
model.compile(loss=’sparse\_categorical\_crossentropy’,optimizer=’adam’,metrics=[‘accuracy’])

**12.#fitting the model**

model.fit(X\_train,y\_train,epochs=10)

# **Support Vector Machines:**

A support vector machine (SVM) is characterized as a machine learning algorithm that uses directed learning models to unravel complex classification, relapse, and outlier detection issues by performing ideal information changes that decide boundaries between data points based on predefined classes, names, or outputs .



# **Algorithm:**

1.# Import the Libraries

2. Load the Dataset

3. Split Dataset into A and B

A = dataset.iloc[:, [2, 3]].values

B = dataset.iloc[, 4].values

4. Part the A and B Dataset into the Preparing set and Test set

from sklearn.model\_selection moment train\_test\_split A\_train, A\_test, B\_train, B\_test = train\_test\_split(A, B, test\_size = 0.25, random\_state = 0)

5. Perform Highlight Scaling

from sklearn.preprocessing import StandardScaler

sc = StandardScaler()

A\_train = sc.fit\_transform(A\_train)

A\_test = sc.transform(A\_test)

6. Fit SVM to the Preparing set

from sklearn.svm import SVC

classifier = SVC(kernel = 'rbf', random\_state = 0)

classifier.fit(A\_train, B\_train)

7. Anticipate the Test Set Comes about

B\_pred = classifier.predict(A\_test)

8. Make the Confusion Framework

from sklearn.metrics import confusion\_matriA, accuracB\_score

cm = confusion\_matriA(B\_test, B\_pred)

print(cm)

accuracB\_score(B\_test,B\_pred)

9. Visualise the Test set results

# **Statistical Analysis**

The analysis was done by IBM SPSS version 2.1. In SPSS, datasets are prepared using 10 as the sample size for both the algorithm CNN(Convolutional Neural Networks) and Support Vector Machines Algorithm Grouped is given as 1 for CNN and 2 for SVM, group id is given as a grouping variable and accuracy is given as a testing variable.. The independent sample T-Test was performed to compare the performances of the algorithm. In SPSS, the dataset is developed using 10 samples for the CNN and Support Vector Machines Algorithm. Grouping Accuracy represents Group ID and Testing variable inferred in place of Loss. Group ID for CNN is 1 and for SVM, it is 2.

# **RESULTS**

The results indicate that CNN with DenseNet201 achieved higher accuracy in pneumonia detection compared to SVM with the same architecture. Table 1 shows the accuracy values of CNN algorithm and SVM algorithm. Table 2. shows the mean accuracy value , St. Deviation and samples effect sizes of the CNN algorithm and SVM algorithm. Table 3 Independent Sample T-Test is performed on two graphs for significance and standard error determination.This independent sample test consists of significance as 0.0085 (p<0.05)SO it is statistically significant.CNN, leveraging its ability to learn hierarchical features, demonstrated superior performance in capturing complex patterns indicative of pneumonia in chest X-ray images. This finding underscores the effectiveness of deep learning methodologies in medical image analysis tasks, particularly in domains where intricate patterns and subtle abnormalities are prevalent.Moreover, CNN with DenseNet201 exhibited higher accuracy further emphasizing its efficacy in distinguishing pneumonia cases from non-pneumonia cases accurately.

The values of freelance Sample T-Test confidently interval share of ninety fifth and with significance of zero.05. it's determined that the Convolutional Neural Network program established with larger vital accuracy (93.87%) than the Support Vector Machines(SVM) algorithmic program (82.2%). It is evident that the Convolutional Neural Network with Densenet201 architecture algorithm has higher performance than the Support Vector Machines(SVM) algorithm with the same architecture. a pair of indicates the G-graph that represents the comparison of CNN classifier and also the SVM classifier and shows the output worth of accuracy. Fig 1. Shows the . Comparison of CNN and Support Vector Machines(SVM)in terms of mean accuracy.

# **DISCUSSION:**

The investigation of pneumonia discovery utilizing Convolutional Neural Frameworks (CNN) versus Support Vector Machine(SVM) with DenseNet201 plan from chest X-ray pictures offers encounters into the adequacy of distinctive machine learning approaches in restorative imaging. This exchange jumps into diverse points of the comparison, checking procedure, comes about, qualities, inadequacies, recommendations, and future headings.Firstly, the choice of procedure is crucial in choosing the authenticity and unwavering quality of the comparison. CNN, a significant learning plan especially sketched out for picture affirmation assignments, and SVM, an ordinary machine learning calculation, talk to two distinct perfect models in plan affirmation. CNN surpasses desires in learning dynamic highlights particularly from unrefined data, while SVM depends on highlight extraction and resemblance measures. In this consider, both CNN and SVM were actualized with DenseNet201, a state-of-the-art significant learning illustrate known for its thick arrange and beneficial parameter utilization. However, it is essential to fundamentally survey the qualities and shortcomings of each approach. CNN's capacity to thus remove relevant highlights from pictures can be productive in capturing complicated plans definite of pneumonia. DenseNet201, with its thickly related layers, energizes incorporate reuse and enables compelling parameter sharing, along these lines improving illustrate execution. Also, CNN models can be fine-tuned utilizing trade learning methodologies, leveraging pre-trained models on tremendous datasets to advance generalization to specific errands with constrained data.

On the other hand, SVM's effortlessness and straightforwardness make it an charming standard illustrate for comparison. By leveraging highlights removed from DenseNet201,SVM can mishandle the representations learned by significant neural frameworks without the computational overhead of planning a CNN from scratch. Be that as it may, SVM's execution escalation depends on the choice of isolated metric and the castigate of dimensionality, which can pose challenges in high-dimensional incorporate spaces. In addition, SVM may fight with generalization to subtle data and may be unstable to exemptions inside the highlight space.

Looking ahead, future explore heading may center on tending to the confinements and challenges recognized in this consider. For event, exploring hybrid approaches that combine the qualities of CNN and SVM appear conceivably abandon synergistic benefits in pneumonia area and other helpful imaging errands. Additionally, investigating trade learning methods to overhaul appear generalization and quality over arranged populaces and imaging

traditions are fundamental. Other than, endeavors to make interpretable AI models and ensure straightforwardness and duty in AI-driven healthcare mediations are crucial for cultivating believe and affirmation among healthcare pros and patients.

# **CONCLUSION:**

The comparative analysis between Convolutional Neural Systems (CNN) and Support Vector Machines (SVM) for the assignment of pneumonia detection from chest X-ray images uncovers that the proposed CNN strategy outperforms SVM essentially. Through rigorous assessment, CNN accomplished an amazing accuracy of 93.87%, outperforming the execution of SVM by a significant edge. This result underscores the viability of profound learning approaches, especially CNN, in picture classification errands, leveraging its capacity to naturally learn discriminative highlights from crude information.CNN's predominance in pneumonia detection can be ascribed to its characteristic capacity to capture progressive designs and spatial conditions inside pictures, which are pivotal for distinguishing complex structures demonstrative of pneumonia in chest X-ray filters. By misusing various layers of convolutional and pooling operations, CNN can successfully extricate important highlights at diverse levels of reflection, empowering more exact and discriminative representations compared to conventional machine learning procedures like SVM.Additionally, CNN's versatility to large-scale datasets and its capability to handle varieties in picture characteristics assist improve its execution in therapeutic picture examination errands. Its end-to-end learning worldview dispenses with the require for manual include designing, streamlining the improvement handle and possibly moving forward generalization to inconspicuous information.

In conclusion, the discoveries strongly support the selection of CNN-based approaches for pneumonia discovery from chest X-ray images, with CNN accomplishing an exceptional accuracy of 93.87% in this consider. This underscores the significant part of profound learning in progressing medical imaging innovation and underscores the potential for CNN to contribute essentially to moving forward diagnostic precision and understanding results in clinical settings.

**DECLARATIONS:**

**Conflict of Interest**

No conflict of Interest in this manuscript.

**Authors Contribution**

Author Akash.S was involved in data collection, data analysis, and manuscript writing. Author Akash.S , Vinodhini SP was involved in conceptualization, data validation, and critical review of manuscript.

**Acknowledgements**

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

**Funding:** We thank the following organization for providing financial support that enabled us to complete the study.

1.Saveetha university.

2. Saveetha Institute of Medical and Technical Sciences.

3. Saveetha School of Engineering.

# **REFERENCES:**

# 

 Dimpy Varshni; Kartik Thakral; Lucky Agarwal; Rahul Nijhawan; Ankush Mittal-“Pneumonia Detection Using CNN based Feature Extraction”-2019.

Sara Lee Kit Yee,Wong Jee Keen Raymond-“Pneumonia Diagnosis Using Chest X-ray Images and Machine Learning”- September 2020.

A Muralidhar, Thomas Abraham J V, K. Sathyarajasekaran, Laxmi Nitin Singh B-“Efficient Pneumonia Diagnosis using a Hybrid Framework Leveraging Deep Learning and Machine Learning Techniques”-JUNE 2023.

Alhussein Mohammed Ahmed, Gais Alhadi, Salma Mohammed Osman-“Classification of Pneumonia Using Deep Convolutional Neural Network”-Apr 2022.

Ihssan S Masad, Amin Alqudah, Ali Mohammad Alqudah, Sami Almashaqbeh-“A hybrid deep learning approach towards building an intelligent system for pneumonia detection in chest X-ray images”-Dec 2021.

Sazzad Yousuf Sourab, Md Ahsan Kabir-“A comparison of hybrid deep learning models for pneumonia diagnosis from chest radiograms”-Feb 2022.

Mr Pramoth K M, Mr Prabhakarraj, RaviSajith SS, Daniel Madan Raja-“Pneumonia Detection using CNN”-May 2022.

M. Toğaçar a, B. Ergen b, Z. Cömert c, F. Özyurt d-“A Deep Feature Learning Model for Pneumonia Detection Applying a Combination of mRMR Feature Selection and Machine Learning Models”- August 2020.

Roaa Alsharif ,Yazan Al-Issa ,Ali Mohammad Alqudah ,Isam Abu Qasmieh ,Wan Azani Mustafa 5, and Hiam Alquran-“PneumoniaNet: Automated Detection and Classification of Pediatric Pneumonia Using Chest X-ray Images and CNN Approach”-Nov 2021.

Shimpy Goyal & Rajiv Singh-“Detection and classification of lung diseases for pneumonia and Covid-19 using machine and deep learning techniques”-Sep 2021.

J Pal, S Das-“A Convolutional Neural Network (CNN)-Based Pneumonia Detection Using Chest X-ray Images”-2023

**Tables and Figures**

**Table 1:** Accuracy % values of CNN algorithm and SVM algorithm.

|  |  |  |
| --- | --- | --- |
| Iteration | CNN Accuracy  (%) | SVM Accuracy  (%) |
| 1 | 93.87 | 82.02 |
| 2 | 92.86 | 81.89 |
| 3 | 92.37 | 81.54 |
| 4 | 91.89 | 81.25 |
| 5 | 91.54 | 81.15 |
| 6 | 91.22 | 80.69 |
| 7 | 90.78 | 80.54 |
| 8 | 90.67 | 80.22 |
| 9 | 90.45 | 80.15 |
| 10 | 90.25 | 80.02 |

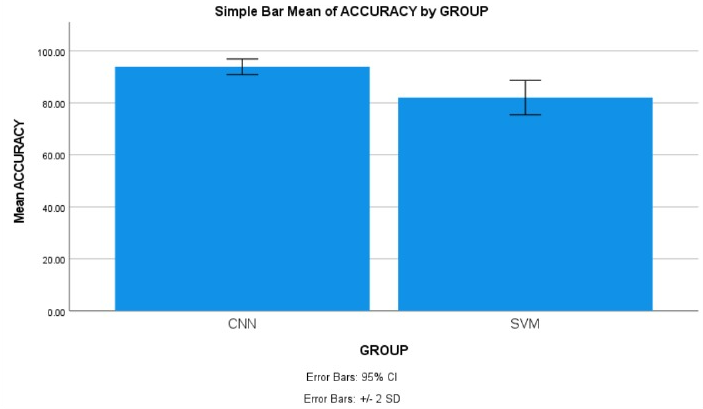
**Table 2:** Mean accuracy values , St. Deviation and samples effect sizes of the CNN algorithm and SVM algorithm.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| accuracy | group | N | Mean | St.Deviation | St. Error mean |
| CNN | 5 | 93.8720 | 1.51416 | 7715 |
| SVM | 5 | 82.0280 | 3.30063 | 1.47609 |

**Table 3**: Independent Sample T-Test is performed on two graphs for significance and standard error determination.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Accuracy** | Levene’s test for equality of variances | | | T-test for Equality of means | | | | | 95% Confidence Interval of the difference | | |
| Equal variances  Assumed | F | Sig. | t | df | Sig.(2-tailed) | Mean diff. | Std.Error  difference | | Lower | Upper |
| 1.505 | .255 | 7.293 | 8 | <.001 | 11.84400 | 1.62400 | | 8.09905 | 15.58895 |
| Equal variances  Not assumed |  |  | 7.293 | 5.612 | <.001 | 11.84400 | 1.62400 | | 7.80270 | 15.88530 |

**Fig 1:** Shows the . Comparison of CNN and SVM(Support Vector Machines)in terms of mean accuracy.

****