Detecting and Digitizing Handwritten Dates in Real Time

Presented By: Lakshmi Sravani

Index

Use case

- In many Hospitals, patients write the date(Signature and date) on a digital pad (e.g., with a stylus).

- The goal of this project is to instantly recognize those handwritten digits and digitize the **date** for accurate and fast medical record-keeping.

Why Handwriting Still Exists in Hospitals

01

Signatures and dates are legally required in many medical forms.

02

Digital pads mimic paper, making the experience intuitive and natural.

03

Patients, especially elderly, feel more comfortable writing than typing.

04

Hybrid systems store handwriting digitally, but don’t interpret it — your system fills that gap.

Image

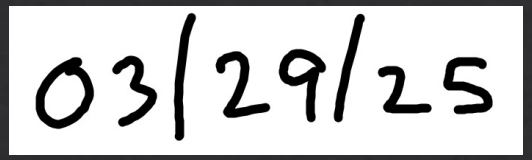
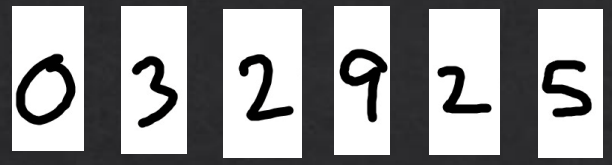


Image of date

Splitting



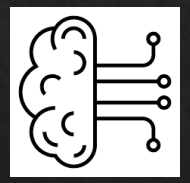
Segmenting the date into individual Digits

Preprocessing

Each digit of size 28*28 pixel (Each digit image has 784 pixels.)

Each array

Model



Naive bayes and Non Naive bayes

Digital character

Combine Digital date

03/29/25

Merge Digital character and form a date

Dataset

- The MNIST dataset is one of the most widely used benchmarks in machine learning for handwritten digit recognition.
- It contains **70,000 labeled images** of handwritten digits from 0 to 9 — all written by real people.
  - **60,000 training images**
  - **10,000 testing images**
- Each image in the dataset is **28×28 pixels**, making a total of **784 features per image**.
- They are **grayscale images**, meaning each pixel holds a value between **0 (black)** and **255 (white)**.

Pre-processing

- Normalize pixel values by dividing by 255.0

  X_train_nor = X_train / 255.0

- Preprocessing ensures all pixel values are in the range [0, 1], improving model performance
- Normalize both train and test dataset
- This setup allows models to focus purely on digit shape, unaffected by intensity variations

Models

Naive Bayes

Naive Bayes assumes pixel independence and computes probabilities using mean and variance. It gave us a decent accuracy of 77.91%.

Non Naive Bayes

Non-Naive Bayes does not assume feature independence, meaning it considers feature correlations by computing the full covariance matrix instead of using only variances. Accuracy of 95.44%

KNN

K-Nearest Neighbors (KNN) is a algorithm that classifies data points based on the majority class of their **K** closest neighbors. Takes long time to run.

Challenges

Segmenting entire handwritten date into individual digits

Recognizing characters like slashes, hyphen (/,-)

Date format can be different i.e., 03/29/2025, 3/29/25, 3-29-2025

Future Work

1

Build a real-time web or tablet interface

2

Automate digit segmentation from date input - CV

3

Expand to recognize different characters slashes, Hyphen and period

Conclusion

We introduced a real-world system to digitize handwritten dates, capable of recognizing various handwriting styles by training the model with 60,000 records.

Using Non-Naive Bayes, we achieve a 95.4% accuracy in digit prediction.

The predicted digits form a date in the MM/DD/YYYY format and are stored in the patient's medical record.

Thank You