Word-Level Pakistani Sign Language Recognition

Authors:

Jahanzeb Naeem

Zain Asif

Arif Ur Rahman

Department of Computer Science,

Bahria University, Islamabad Campus

Contents

Introduction

01 Over 430 million people suffer from hearing loss; 70 million use sign language for communication

02 In Pakistan, 250,000+ rely on PSL for daily communication

03 Existing PSL systems face two key challenges: reliance on static images and limited to basic alphabets

Introduction

04 Proposed a video-based word-level PSL dataset covering 31 classes

05 Developed a recognition system using customized RNN architectures (LSTM, GRU, BRNN)

06 Achieved 97% validation accuracy with LSTM

Motivation

- Current PSL recognition systems do not handle word-level recognition due to limited datasets and static data
- Sign languages inherently involve motion; capturing these temporal dynamics is essential for accurate recognition
- Addressing this gap can greatly improve accessibility for the deaf-mute community

Problem

- Incomplete datasets covering only alphabets
- Static image-based datasets fail to capture body movements
- Variability in PSL signs across different regions

Solution: A comprehensive, motion-aware PSL dataset for word-level recognition

Literature Review

- Prior work focused mainly on American Sign Language (ASL) datasets (e.g., WLASL)
- PSL datasets limited to static images or alphabets; previous efforts include Mirza et al. and Javaid et al.
- Initially relied on Kmeans and CNN, now moving towards bigger models (SLATN)
- Pose estimation approach missing

Methodology

1. Data Collection and Annotation

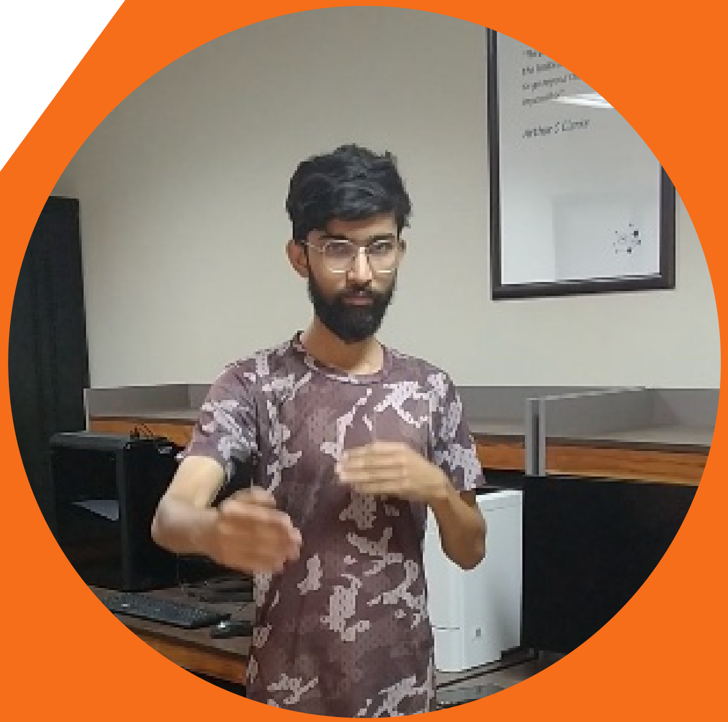
- Collected data for 31 PSL word classes, with 248 video samples representing varied environmental conditions (indoors, outdoors, different lighting)
- Ensured video quality consistency while capturing different backgrounds, clothing, and lighting scenarios for diverse and high-applicability training
- All videos belonging to the same sign were stored in a separate directory with the name of the directory serving as the label

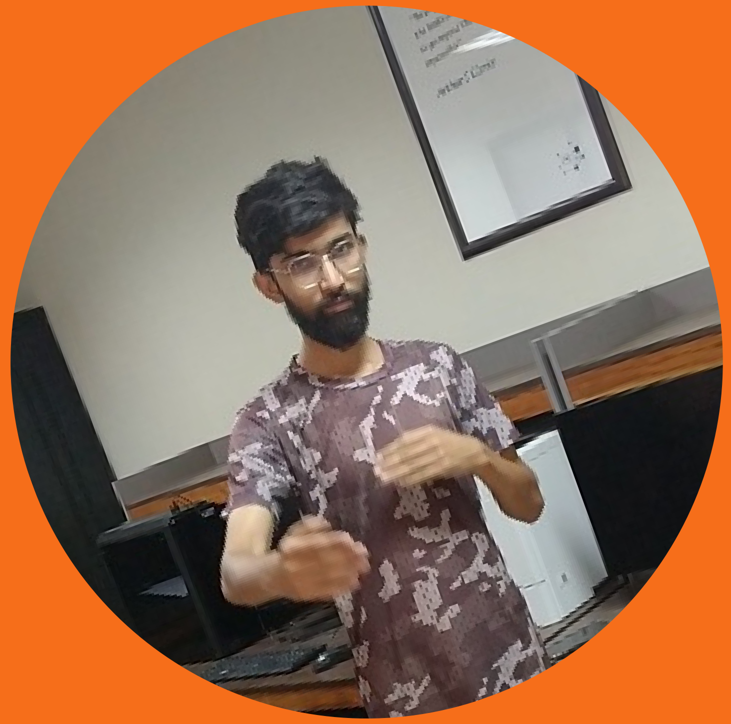
Methodology

2. Data Processing

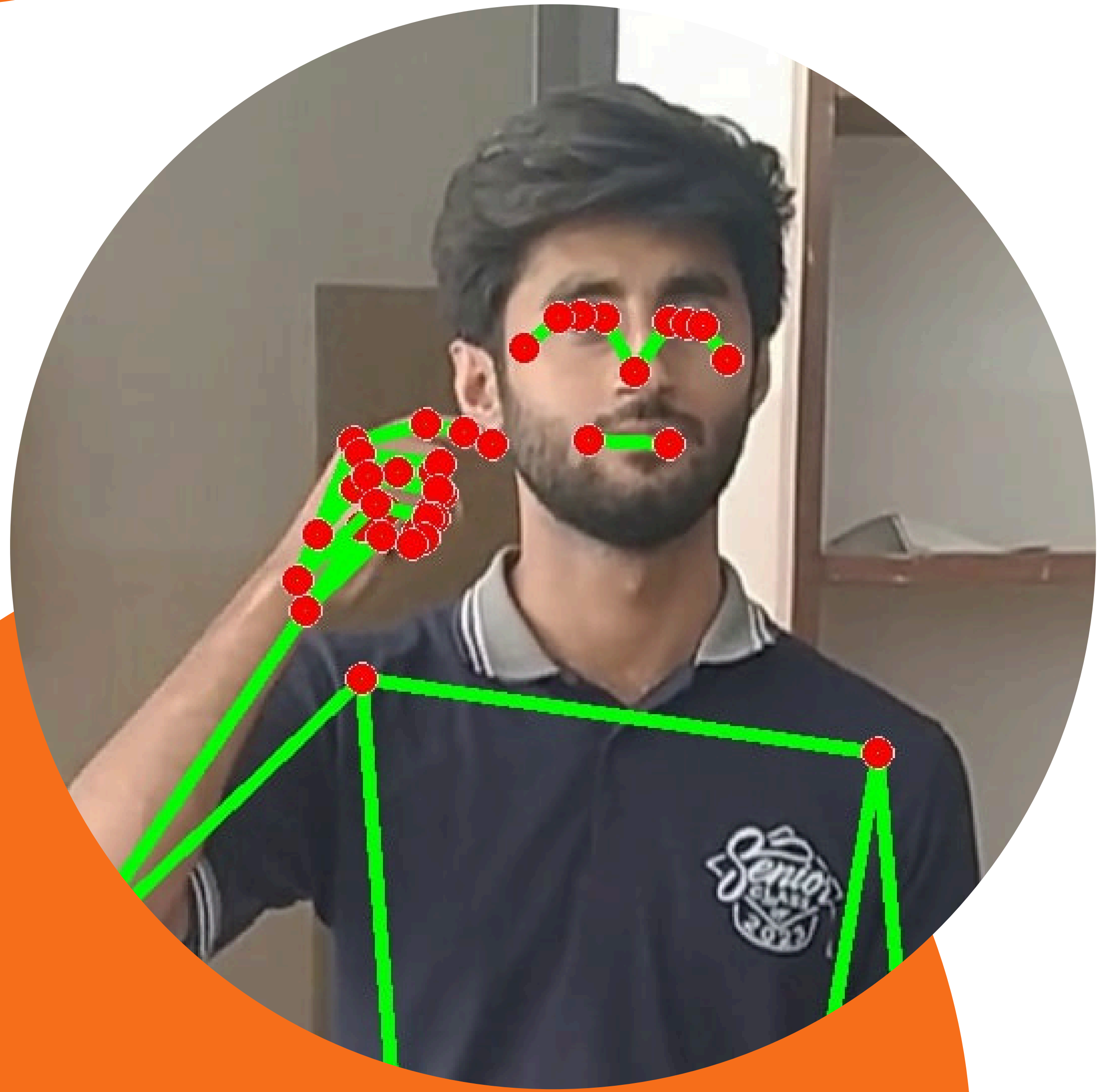
- Selected several frames from each video to capture important gestures

- Rotations, flips, brightness changes applied to simulate various real-world scenarios and enhance dataset diversity





Methodology



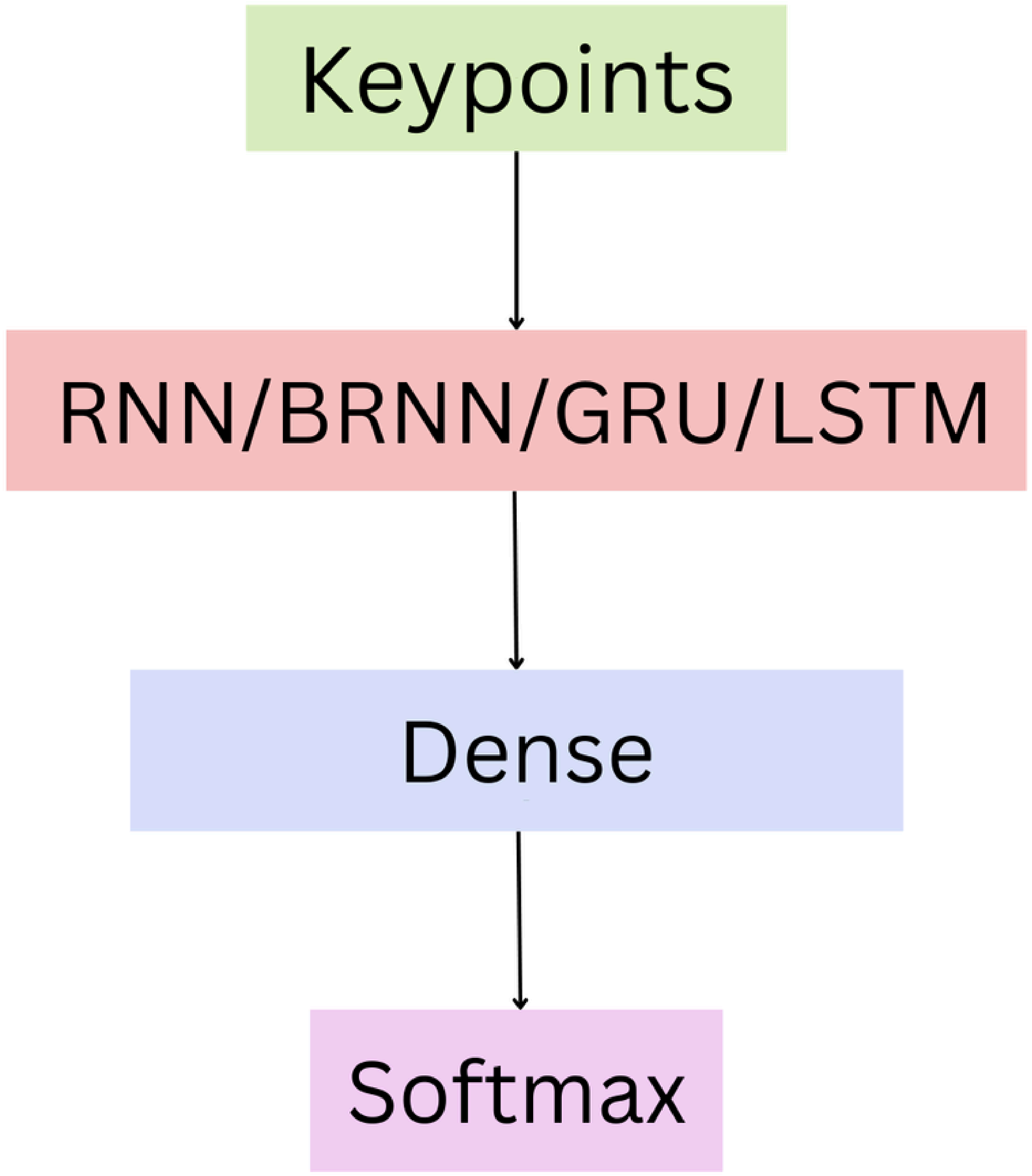
3. Keypoints Extraction

01 Captured 3D spatial coordinates for each keypoint frame-by-frame to model movement patterns

02 Google Mediapipe library's Holistic model used

03 258 keypoints per frame, 50 frames per video, all concatenated into one big 3D array

Methodology



4. Model Training

01 Explored multiple RNN-based approaches: Basic RNN, Bidirectional RNN (BRNN), Gated Recurrent Unit (GRU), and Long Short-Term Memory (LSTM)

02 Aimed to capture sequential dependencies in gesture movements with different model architectures

03 Evaluated models based on validation accuracy and loss metrics

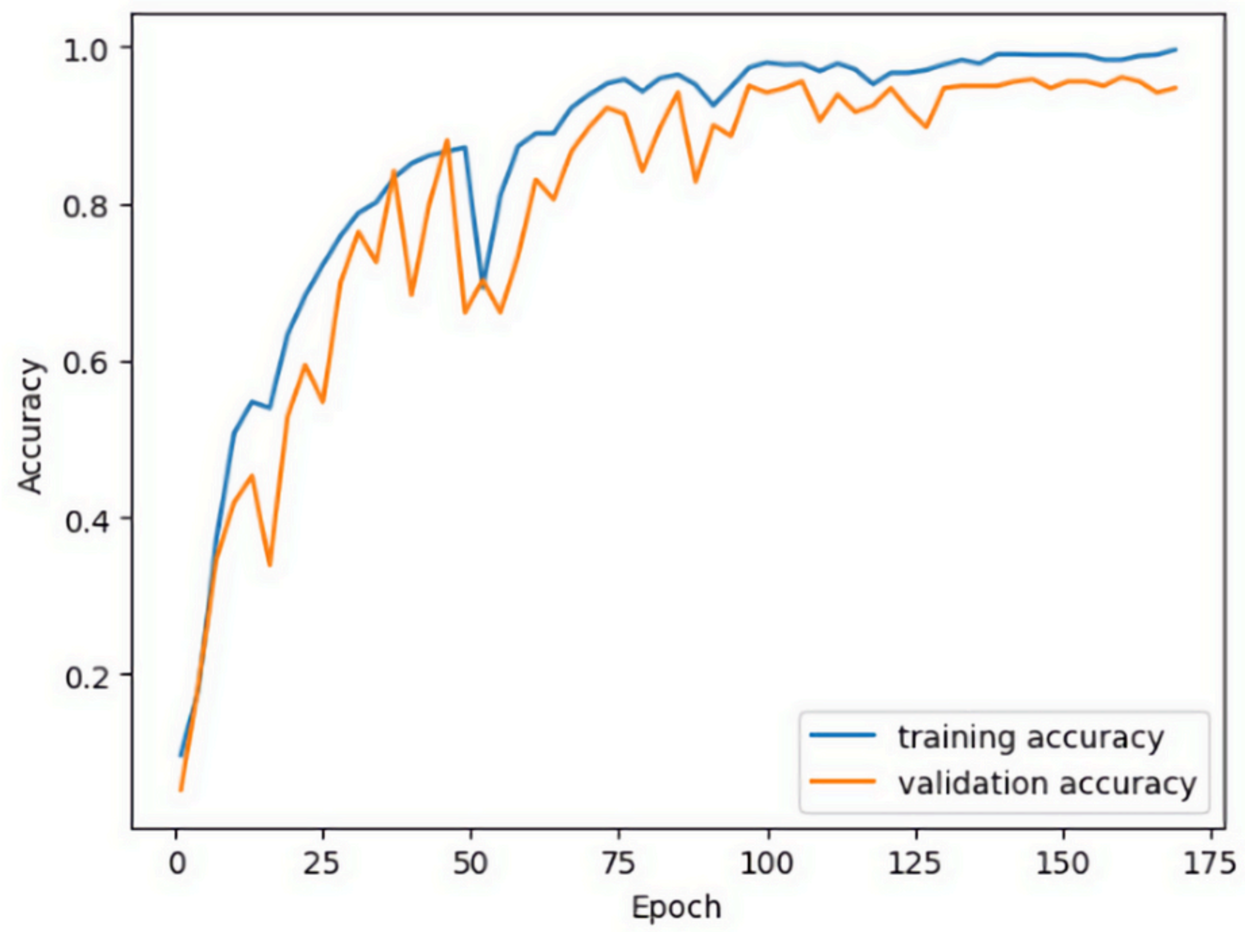
Results

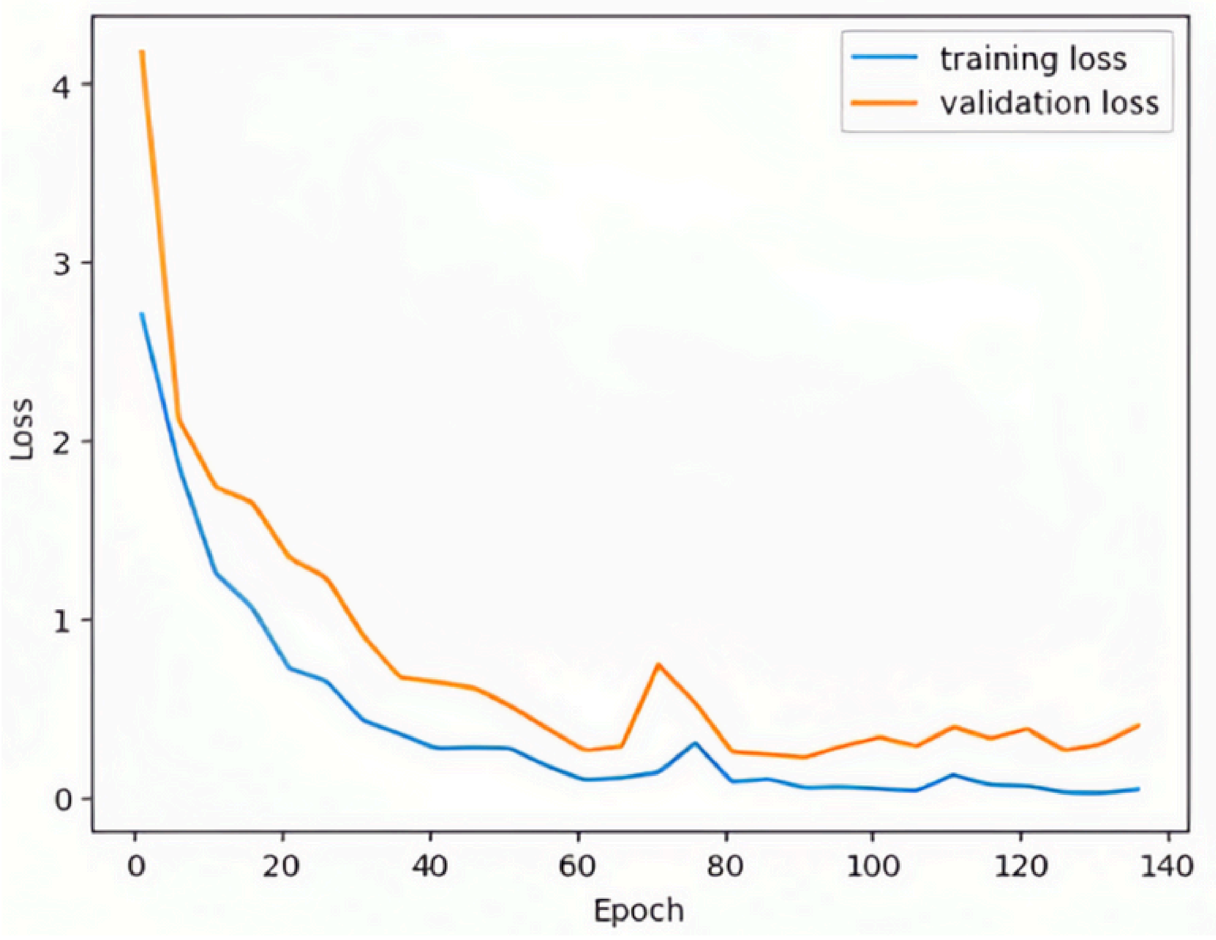
01 LSTM achieved the highest accuracy (97%) for the 15-class problem

02 Trained using early stopping and learning rate adjustments to optimize performance

03 Loss reduced significantly during the first 20 epochs, converging around 0.2

04 Comparison with other models: GRU (93%), BRNN (89%), RNN (85%)





Demo Videos

1

2

Analysis

01 RNN-architecture succeeded due to its ability to capture long-term dependencies in sign language sequences

02 BRNN struggled early on, but after 30 epochs, performance greatly improved

03 LSTM and GRU ended up pretty close to each other with LSTM slightly ahead

04 GRU offers competitive accuracy and faster convergence

Future Work

With this project, our main aim was to help bring the attention of researchers towards the major gaps in this field of research. We hope that one day our dataset can help researchers build a system that can detect and translate Pakistani Sign Language reliably and in real-time.

Conclusion

Communication is an essential part of our lives and it is our responsibility as computer scientists to help those struggling to fulfill this need

This dataset is a small step in that direction and with continued support, it has the potential to have an impact on millions of lives

Thank You

Contact:
naeemjahanzeb@gmail.com
zainasifsatti@gmail.com
arif.buic@bahria.edu.pk