Universidad de Málaga

Escuela de Doctorado

Escuela Técnica Superior de Ingeniería de Telecomunicación

Programa de Doctorado en Ingeniería de Telecomunicación

TESIS DOCTORAL

Machine Learning for Bidirectional Translation between Different Sign and Oral Languages

Autor:

Muhammad Imran Saleem

Director:

Miguel-Ángel Luque-Nieto

Málaga, 2023

AUTOR: Muhammad Imran SALEEM

https://orcid.org/0000-0001-9138-9290

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga

UNIVERSIDAD DE MÁLAGA

Escuela de Doctorado
Edificio Pabellón de Gobierno
29071 Málaga
952 131 028 / 952 131 461 / 952 137 110
doctorado@uma.es

DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA PARA OBTENER EL TÍTULO DE DOCTOR

Muhammad Imran SALEEM

Alumno del programa de doctorado en INGENIERÍA DE TELECOMUNICACIÓN de laEscuela de Doctorado de la Universidad de Málaga, autor de la tesis, presentada para laobtención del título de doctor por la Universidad de Málaga, titulada: *Machine learning for bidirectional translation between different sign and oral languages*,

Realizada bajo la tutoría de Pablo Otero y la dirección de Miguel Ángel Luque Nieto.

DECLARO QUE:

La tesis presentada es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, conforme al ordenamiento jurídico vigente (Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), modificado por la Ley 2/2019, de 1 de marzo.

Igualmente asumo, ante a la Universidad de Málaga y ante cualquier otra instancia, la responsabilidad que pudiera derivarse en caso de plagio de contenidos en la tesis presentada, conforme al ordenamiento jurídico vigente.

En Málaga, a 19 de junio de 2023

Imran Saleem Fdo.: Muhammad Imran Saleem Alumno de Doctorado	P. Otero Fdo.: Pablo Otero Roth Tutor
Fdo.: Miguel Ángel Luque Nieto Director de tesis	

This page intentionally left blank

UNIVERSIDAD
DE MÁLAGA

AUTORIZACIÓN DE LOS DIRECTORES DE TESIS DOCTORAL

El alumno del programa de doctorado en Ingeniería de Telecomunicación, Muhammad Imran SALEEM, es primer autor de las siguientes publicaciones en revistas indexadas en los *Journal CitationReports* (JCR) del *Web of Science* (WoS):

- **Saleem, M.I.**; Siddiqui, A.; Noor, S.; Luque-Nieto, M.-A.; Otero, P. A Novel Machine Learning Based Two-Way Communication System for Deaf and Mute. *Appl. Sci.* **2023**, *13*, 453. https://doi.org/10.3390/app13010453 (JCR-Q2).
- **Saleem, M.I.**; Siddiqui, A.; Noor, S.; Luque-Nieto, M.-A.; Nava-Baro, E. A Machine Learning Based Full Duplex System Supporting Multiple Sign Languages for the Deaf and Mute. *Appl. Sci.* **2023**, *13*, 3114. https://doi.org/10.3390/app13053114 (JCR-Q2).

Estas publicaciones avalan su tesis doctoral y ninguna otra tesis.

Por todo ello, su tutor Pablo Otero Roth y su director de tesis Miguel-Ángel Luque-Nieto autorizan al Sr. Muhammad Imran Saleem a depositar su tesis doctoral ante las Autoridades Académicas de la Universidad de Málaga.

En Málaga, a 19 de junio de 2023

P. Otero

Fdo: Pablo Otero

Fdo: Miguel Ángel Luque Nieto

This page intentionally left blank

Abstract

Deaf and mute people are an integral part of society, and it is particularly important to provide them with a platform to be able to communicate without the need for any training or learning. These deaf and mute (D-M) individuals, who rely on sign language, but for effective communication, it is expected that others can understand sign language. Learning sign language is a challenge for those with no impairment. In practice, D-M face communication difficulties mainly because others, who generally do not know sign language, are unable to communicate with them. The work presents a solution to this problem through a system enabling the non-deaf and mute (ND-M) to communicate with the D-M individuals without the need to learn sign language. Another challenge is to have a system in which hand gestures of different languages are supported. In this thesis, a system ispresented that provides communication between D-M and non-deaf and mute (ND-M). The hand gestures of D-M people are acquired and processed using deep learning, and multiple language support is achieved using supervised machine learning. The D-M people are provided with a video interface where the hand gestures are acquired. and an audio interface to convert the gestures into speech. Speech from ND-M people is acquired and converted into text and hand gesture images. The system is easy to use, low cost, reliable, modular, based on a commercial-off-the-shelf (COTS) Leap Motion Device (LMD), and can be enhanced by adding data to support more languages. A supervised machine learning dataset is created that provides multi-language communication between the D-Mand ND-M people. This system will support D-M people in communicating effectively with othersand restoring normalcy in their daily lives.

A supervised ML algorithm, using a Convolutional Neural Network (CNN), converts the hand gesture data into speech. A new dataset for the ML-based algorithm is created and presented in this manuscript. This dataset includes three sign language datasets, i.e., American Sign Language (ASL), Pakistani Sign Language (PSL), and Spanish Sign Language (SSL). The proposed system automatically detects the sign language and converts it into an audio message for the ND-M. Similarities between the three sign languages are also explored, and further research can be carriedout in order to help create more datasets, which can be a combination of multiple sign languages. The system also provides a training mode that can help D-M individuals improve their hand gestures and also understand how accurately the system is detecting these gestures.

The proposed system has been validated through a series of experiments. The hand gesturedetection accuracy of the system is more than 90% for most, while for certain scenarios, this is between 80% and 90% due to variations in hand gestures between D-M people.

Resumen

Introducción

Las personas sordas o con discapacidad auditiva son parte de una sociedad en la que la mayoría son oyentes. Esta diferencia da lugar a que el riesgo de exclusión social de las personas sordas sea altísimo. El principio de igualdad y la idea de dignidad constituyen una responsabilidad y unaobligación para las personas oyentes y sus gobernantes de abordar el problema de la exclusión. La principal herramienta para la comunicación de las personas sordas entre sí es la lengua de signos. No obstante, para una comunicación y una inclusión social efectivas, se espera que otras personas, los denominados hablantes, también puedan entender la lengua de signos y expresarse en ella. Sinembargo, aprender la lengua de signos es un desafío para aquellas personas sin esa discapacidad. La realidad es que muy pocas personas oyentes poseen apenas un mínimo conocimiento de la lengua de signos de su país o región. Este hecho es una condena a la incomunicación para las personas sordas, con las graves consecuencias de aislamiento y de exclusión social de las personas sordas, lo que puedeconducir incluso a patologías severas. Este problema se agrava si se tiene en cuenta que, al igual que ocurre con las lenguas habladas, los distintos idiomas tienen diferentes lenguas de signos.

La Organización Mundial de la Salud (OMS) informa de que el 5% de la población mundial son personas sordas o con discapacidad auditiva. Si el dato es correcto, eso significa que son más de 400 millones de personas en el mundo (recientemente la población mundial ha superado la cifra de 9.000 millones de personas).

Una inmensa mayoría de países del mundo han reconocido el derecho fundamental de las personas sordas a la comunicación y han actuado en consecuencia, al menos desde el punto de vista de su legislación nacional. Por ejemplo, en España la Ley 27/2007, de 23 de octubre, reconoce las distintas lenguas de signos españolas y regula los medios de apoyo a la comunicación oral de las personas sordas, con discapacidad auditiva y sordociegas. Otra cosa es su desarrollo y su cumplimiento, pero hay que decir que se avanza en esa dirección.

Los medios de apoyo mencionados en el título de la Ley son diversos, desde la asistencia personal, mediante traductores de lengua de signos o las oficinas de asistencia social de Ayuntamientos, hasta los medios quirúrgicos como son los implantes cocleares. Quizá los implantes cocleares, que se llevana cabo a temprana edad del niño con discapacidad auditiva, lleguen algún día a hacer innecesarias las lenguas de signos. O incluso otro avance de la combinación medicina/ingeniería que aún no podemos imaginar. Pero la realidad actual es que la lengua de signos será por muchos años el principal medio de comunicación de estas personas.

Entre ambos extremos de los medios de apoyo se encuentra la tecnología a secas, sin intervención de la medicina. Parece lo más natural que las tecnologías asociadas a la ingeniería de telecomunicación, como son la electrónica de sonido e imagen, el procesado de voz y de imagen, la transmisión de datos o los algoritmos que dieron origen a lo que hoy en día se conoce como inteligencia artificial, proporcionen medios que conviertan voz en imágenes (de signos) y viceversa.

A modo de sumario, en la presente tesis doctoral se presenta un sistema que proporciona comunicación entre personas sordas (D-M, del inglés Deaf-Mute, usado en toda la Memoria de tesis)

y oyentes (ND-M, de non Deaf-Mute, también en el resto de la Memoria). Los signos de la lengua de signos son gestos realizados con una o dos manos. El sistema propuesto adquiere y procesa los signos utilizando el aprendizaje profundo (*Deep Learning*). El soporte multi-idioma se consigue utilizando aprendizaje automático supervisado (*Supervised Machine Learning*).

En lo que respecta a la operativa del sistema, para entender a las personas sordas, las personas oyentes cuentan con una interfaz vídeo/audio donde los signos de la mano se convierten en habla. El sonido lo produce la tarjeta de audio de un ordenador/tableta. En sentido contrario, el habla de las personas ND-M es registrado por la misma tarjeta de audio y convertido en texto mediante un reconocedor de voz. El sistema propuesto es fácil de usar, de bajo coste, modular y puede ser mejorado añadiendo datos para soportar más idiomas en el futuro.

En cuanto al interior del sistema, en esta tesis se ha definido y creado una base de datos de aprendizaje automático supervisado que proporciona comunicación multilingüe automática entre las personas sordas y las hablantes. El sistema ayuda a las personas de D-M a comunicarse de manera efectiva con los demás y a restaurar una sensación de normalidad en su vida diaria.

Para completar este breve sumario es importante mencionar que se han llevado a cabo experimentos para determinar la eficacia o precisión de detección de gestos de la mano del sistema, que ha resultado ser superior al 90% para la mayoría de los escenarios. En ocasiones, para ciertos escenarios que se comentan en la memoria, la precisión ha quedado entre el 80% y el 90% debido a variaciones en los gestos en la mano de unas personas sordas a otras. En el mundo de los sordos también la lengua de signos admite adjetivos como acento, firmeza, arrogancia, dulzura, etc. Esas características pueden disminuir la precisión del sistema, pero este inconveniente puede solventarse invitando a los usuarios a que sean más asépticos en la expresión, lo que equivale a pedirle a un hablante que vocalice.

Organización de la Memoria de Tesis

La memoria de tesis se ha organizado en cinco capítulos, siguiendo un patrón convencional de este tipo de memorias. En el primer capítulo se presentan el escenario al que se ha enfrentado esta investigación. Además, se han reunido las principales características de los sistemas y tecnologías que intervienen en el escenario, a saber, la lengua de signos, el *Machine Learning* (ML), el *Deep Learning* (DL), las redes neuronales artificiales y las redes neuronales convolucionales, que son los elementos sobre los que se ha basado la investigación.

El capítulo 2 es un resumen del estado del arte en las tecnologías que se han desarrollado previamente para alcanzar el objetivo de la Ley 27/2007, nombrados allí como medios de apoyo. Además, en el capítulo se exponen los últimos avances en ML y DL.

Los capítulos 3 y 4 constituyen el núcleo de la investigación. En la práctica, son extensiones de los artículos publicados que avalan la tesis. El capítulo 3 describe un sistema bidireccional de comunicación entre personas sordas que usan lengua de signos y personas oyentes y hablantes. Además de describir el sistema, se exponen la metodología seguida, las características del sistema quelo hacen diferente de otros, se completa con la descripción de un procedimiento de evaluación del sistema y de su eficacia y, por último, se da una cifra de mérito del comportamiento del sistema.

En el capítulo 4 se describen qué modificaciones y ampliaciones pueden hacerse para que el sistema sea multi-lengua y se siguen los mismos pasos metodológicos del capítulo anterior.

En el capítulo 5 se exponen las principales conclusiones obtenidas de los trabajos de investigación y se propone alguna línea futura que pueda servir de inspiración a otros investigadores que puedan sentir interés por esta línea de investigación.

Descripción del sistema

El sistema propuesto proporciona comunicación entre personas sordas y sordomudas (D-M) y oyentes/hablantes (ND-M). Las personas D-M signan, es decir, se expresan mediante lenguaje de signos. Aunque el sistema queda mejor descrito mediante los diagramas de bloques de hardware y software en el manuscrito, se hará aquí un breve resumen del mismo.

Un dispositivo *Leap Motion Device* (LMD) recoge los signos y los entrega a un ordenador. Un LMD consiste básicamente en un conjunto de luces y cámaras infrarrojas que captan movimientos de objetos situados a corta distancia, junto con un procesador que codifica esos movimientos. Normalmente cuentan con un interfaz USB para entregar esos códigos a un ordenador. Hoy en día, hay LMDs disponibles comercialmente (Commercial Off-The-Shelf, COTS) lo que significa que la disponibilidad del dispositivo no es una limitación del sistema.

El sistema propuesto los procesa mediante un algoritmo de Red Neural Convolucional (CNN). Un algoritmo de *Deep Learning* (DL) identifica los signos. Una vez procesados e identificados, los signos se convierten a texto y a voz. El texto puede leerse en la pantalla y, si el equipo dispone de un interfaz de sonido, como hoy en día es cierto en cualquier dispositivo electrónico como un ordenador personal, una tableta o un teléfono móvil, lo convierte en habla.

En el caso de distintas lenguas de signos, el sistema detecta automáticamente mediante un algoritmo de *Machine Learning* (ML). Para ello, se han creado tres bases de datos, para tres lenguas de signos distintas: la pakistaní (*Pakistani Sign Language*, PSL), por ser la lengua del autor de la tesis, la española (SSL), por ser la hablada en la Universidad donde se ha llevado a cabo esta investigación yla estadounidense (ASL), por ser de facto la lengua franca contemporánea.

El sistema está preparado para incorporar nuevas lenguas de signos, si se crean los correspondientes bases de datos que necesita el algoritmo ML para su identificación.

Esta característica multi-lengua permitirá en el futuro explorar similitudes entre los distintos lenguajes de signos, lo que sin duda ayudará al desarrollo de la lengua internacional de signos, conocida como Sistema de Signos Internacional (SSI), que ahora tiene escasa difusión.

Para el sentido contrario, es decir la comunicación del (ND-M) al D-M) se realiza mediante la grabación sonora convencional de la voz del hablante. A continuación, un reconocedor de voz lo convierte en texto y, lo que es nuevo, ese texto se convierte en signos en la pantalla del ordenador.

Una funcionalidad de la que se ha dotado al sistema, como se ha mencionado más arriba, es que hay un modo de entrenamiento para ayudar a las personas D-M a realizar los signos de forma que el reconocimiento sea más exacto. Como resultado numérico, se puede señalar que en el caso de una persona D-M entrenada de esta manera, la identificación del signo es correcta al 95%.

En personas D-M no entrenadas, esa cifra de mérito puede llegar a reducirse al 80%, aunque una mayoría de personas D-M consiguen alrededor del 90%.

La obtención de estas cifras se llevó a cabo mediante experimentos planificados y ejecutados como se describe en la Memoria, que sirvieron para validar el concepto y estimar su eficacia. Como se ha mencionado, la cifra de mérito utilizada es el acierto en la identificación del signo realizado por una persona D-M.

El sistema propuesto es de bajo coste, modular, fácil de instalar y fácil de usar. Su principal inconveniente es que necesita de un LMD y de un ordenador, lo que en la actualidad está al alcance de cada vez más personas y, por descontado, pueden dotarse varias unidades en lugares como escuelas, hospitales, oficinas de la administración, oficinas de información, etc., es decir, lugares donde deben interactuar personas D-M con personas ND-M que no saben signar.

Características del sistema y operación

Las seis principales características del sistema propuesto son:

- Sistema basado en ordenador personal.
- Uso de dispositivo tipo *Leap Motion*.
- Comunicación bidireccional.
- Detección de distintas lenguas de signos.
- Creación bases de datos de imágenes en distintas lenguas.
- Uso de algoritmos tipo CNN, ML y DL para entrenamiento y reconocimiento.

Hay una opción de capacitación disponible, que permite personalizar las características para los usuarios individuales. El algoritmo basado en ML proporciona una opción de mejora continua donde los datos de nueva y más alta calidad, es decir, las imágenes de los signos realizados con la mano pueden reemplazar una imagen existente.

El sistema propuesto soporta las lenguas de signos estadounidense (ASL), paquistaní (PSL) y española (SSL). Se pueden añadir más lenguas.

El uso del dispositivo LMD es preferible a la utilización de otro tipo de dispositivo, como los de tipo guante equipado con sensores de movimiento, porque no requiere mantenimiento o calibración.

El sistema también almacena los datos adquiridos y procesados y los utiliza para su funcionamiento. Debido a esta característica, la precisión del sistema aumenta debido a que se añaden más datos a la base de datos. El sistema proporciona un modo de entrenamiento del usuario (no confundir con el entrenamiento de los algoritmos, que se describe en un epígrafe posterior), donde el usuario puede comprobar la exactitud del sistema mientras se entrena en su uso. En este modo, se procesan los gestos de la mano (signos) de la persona D-M, y se muestran en texto los resultados detectados. La persona D-M puede variar la forma de signar e ir comprobando la precisión de la detección. En este modo, el sistema también actualiza la base de datos reemplazando las imágenes existentes conimágenes de mayor calidad. El proceso basado en ML compara las imágenes recién adquiridas con las imágenes almacenadas y luego decide entre reemplazar la imagen existente con la nueva o almacenar la nueva imagen junto con la existente o simplemente no guardar la imagen nueva. Tener másarchivos de imagen significa una base de datos mejorada que aumentará la precisión, pero esto también significa que el sistema requerirá más tiempo de procesamiento. Es importante tener un equilibrio entre estos parámetros. Esta decisión es tomada por la implementación basada en ML.

El algoritmo ML también revisa los datos almacenados por el usuario, lo que ayuda a aumentar la velocidad y la precisión del procesamiento. Por ejemplo, si un perfil de usuario muestra que el usuario solo entiende SSL, el proceso ocultará el paso de detección de idioma para ese usuario en particular. Del mismo modo, el usuario puede actualizar el perfil añadiendo más conjuntos de datos de lenguaje de signos y otros parámetros. El sistema también mantiene un registro de rendimiento, tanto para el usuario individual como para el sistema en general, utilizando los datos de usuario almacenados y nuevos datos adquiridos a través del entrenamiento y los modos normales.

El sistema también mantiene un registro de rendimiento, tanto para el usuario individual como para el sistema en general, utilizando los datos de usuario almacenados y nuevos datos adquiridos a través del entrenamiento y los modos normales.

Teniendo en cuenta lo alentadores que han sido los resultados y que se usan dispositivos COTS como son un ordenador personal y un LMD, puede convertirse un prototipo en un producto sin grandes costes de industrialización (diseño, fabricación, integración y pruebas). Otra ventaja es que no se requiere calibración en comparación con otros dispositivos personalizados, que normalmente se basan en sensores conectados a un guante y requieren mantenimiento continuo para asegurarse de que los sensores funcionen correctamente. La actualización de dispositivos personalizados también es un reto donde la disponibilidad de componentes ha sido un problema últimamente.

Otra característica del sistema propuesto es que no es necesario que el usuario final se someta a entrenamiento riguroso. La facilidad de uso es una característica clave de los propósitos esta investigación. Otra característica clave es que los individuos de D-M también pueden usar este sistema para comunicarse entre sí. En este modo, el sistema adquirirá los datos, lo que ayudará a mejorar el rendimiento del algoritmo de aprendizaje automático. La aplicación de software es fácil de instalar y no requiere ninguna licencia. El software procesa los nuevos datos y los almacena en la basede datos.

Por último, hay que mencionar que la aplicación de software para el ordenador está desarrollada en entorno LabVIEW para sistema operativo Windows. Esta elección permite una interfaz gráfica de usuario de alta calidad (GUI) y desarrollo rápido. En el futuro, la interfaz gráfica podría ser modificada para agregar más características y aumentar por tanto las prestaciones del sistema. Por ejemplo, se pueden agregar gestos de varios idiomas a la base de datos para soportar el soporte multilingüe. El reto será procesar gestos similares de diferentes idiomas.

Metodología y validación

La metodología de investigación seguida se organiza en cuatro etapas: concepto y análisis de alternativas (se han analizado las limitaciones de los sistemas existentes), desarrollo y algoritmos, entrenamiento y validación y análisis crítico. En esta sección del resumen trataremos del entrenamiento y la validación, porque la consideramos la etapa más crítica y más significativa desdeel punto de vista metodológico.

El uso de una base de datos conocida es la clave para entrenar un sistema basado en ML, porque constituye la información de la que la "máquina aprende". Un sistema plenamente operativo yvalidado abrirá la puerta a la investigación adicional al procesar nuevas bases de datos a través del sistema.

La selección de ASL, PSL y SSL se ha basado en la disponibilidad, en el tamaño de las bases de datos y en el número de personas que utilizan las respectivas lenguas. Una base de datos pequeña significa que el tamaño del conjunto de entrenamiento también será pequeño; por lo tanto, el sistema entrenado será menos eficaz o exacto, mientras que una base de datos grande resultará en una mayor precisión pero una respuesta más lenta.

Para entrenar el sistema propuesto, se ha utilizado la lengua de signos estadounidense, ASL, debido a la disponibilidad de una base de datos completa y de un tamaño que se ha considerado adecuado, lo que facilita el trabajo de entrenamiento de los algoritmos. Además, permite la comparación del sistema propuesto con otros sistemas en el futuro. Además, por tratarse de la lengua de signos de uso más extendido, en el sentido de número de personas que la conocen y la utilizan, la validación puede ser estadísticamente más consistente.

No obstante, una revisión exhaustiva de la base de datos de la lengua de signos con el objetivo de crear una nueva base de datos está fuera del alcance de lo que se presenta en este manuscrito. Lo que no es óbice para que puedan considerarse otros parámetros para que la base de datos puede ser modificada y redistribuida, en código abierto, siempre que use su propia gramática única, por cuántas personas la están usando.

Para cualquier nuevo sistema propuesto, el primer paso es siempre seleccionar los datos de prueba de entrada conocidos y determinar la salida esperada. A través de este enfoque, el sistema puede ser validado, las limitaciones pueden ser comprendidas y la precisión puede ser definida. Una vez hecho esto, el segundo paso es expandir el conjunto de datos de prueba de entrada y procesarlo a través del sistema. Este proceso ayudará a mejorar las capacidades de este nuevo sistema.

Aquí se define una estrategia de entrenamiento en la que se fijan algunos parámetros para evaluar el sistema propuesto. El primero es seleccionar un conjunto de datos para gestos de mano. Para ello, se utilizan las imágenes de Kaggle[1]. Como se ha mencionado, las imágenes de signos de una mano utilizadas para el entrenamiento y la validación del sistema propuesto son de ASL. Este proceso se utiliza para revisar el sistema propuesto para la precisión, velocidad, facilidad de comunicación y eficacia del algoritmo de aprendizaje automático. Las imágenes utilizadas son en color, formato RGB, aunque también se pueden utilizar otros formatos.

Los alfabetos A a Z y los números 0 a 9 se utilizan para la validación. Dentro de la base de datos hay 1820 imágenes de alfabeto A a Z y 700 imágenes de números 0 a 9. Estas imágenes muestran diferentes variantes de estos alfabetos y números. Este número de imágenes es adecuado para la evaluación del algoritmo.

El conjunto de datos se divide en dos grupos. El primero incluye el 70% de los datos, que se utiliza para el entrenamiento, mientras que el 30% restante se usa para las pruebas (validación cruzada

[1] Kaggle Dataset. Available online: https://www.kaggle.com/datasets/alexalex1211/aslamerican-sign-language (accessed on 23 October 2022).

70/30). Esta es una división estándar para evaluar algoritmos basados en aprendizaje automático. En otras técnicas, los investigadores utilizan el 60% de los datos para la formación, mientras que el 20% cada uno se utiliza para la validación y la prueba, respectivamente.

La validación cruzada se hace separadamente para el alfabeto y para los números. Para la clasificación alfabética, 1274 imágenes, es decir, el 70% de 1820, se utilizan para el entrenamiento, mientras que el 30% restante, o sea, 546 imágenes, se usan para las pruebas. Del mismo modo, para la evaluación de números, el 70% de las imágenes, es decir, 490 de las 700, se asignan a entrenamiento, mientras que las restantes 210 imágenes se utilizan para las pruebas.

Finalmente, el sistema es validado a través de una serie de experimentos, y las limitaciones de algunas características también se han definido. Los experimentos persiguen determinar la exactitud de la detección de los gestos de la mano (signos), el procesamiento de los signos, la evaluación de los signos con calidad de imagen variable, la detección de variaciones en los mismos signos y la identificación y el procesado de diferentes signos que parecen similares, la detección de los signos con presencia de otros objetos visibles como pueden ser relojes, anillos, pulseras, etc. y finalmente la identificación automática de un signo perteneciente a la base de datos de la lengua de signos.

Los experimentos llevados a cabo para la validación y la determinación de la tasa de éxito han sido los siguientes:

- Experimento 1. Exactitud: se ha definido la cifra de exactitud como el cociente entre el número de aciertos sobre el número total de muestras. Se introducen al sistema las imágenes de los signos y el sistema decide cuál es el signo. Se han usado únicamente letras y números que puedan signarse con una sola mano. La cifra de exactitud se muestra en la tabla 3 de la Memoria.
- Experimento 2. Exactitud en secuencia: similar al anterior, pero usando secuencias de tres signos.
- Experimento 3. Mejora de la cifra de aciertos mediante el procesado de la imagen: se trataba de evaluar qué técnicas de procesado de imagen daban lugar a un incremento de la tasa de acierto.
- Experimento 4. Calidad de la ejecución del signo: qué particularidades en la ejecución del signo (que es el equivalente a lo que en audio se llama inteligibilidad, que es una combinación de articulación, entonación, acento, etc.) produce también una mejora.
- Experimento 5. Discriminación entre signos parecidos: dos signos distintos pero que, por ejemplo, consistan en colocar uno o dos dedos en una posición similar, se confunden fácilmente. Este experimento midió la tasa de éxito entre signos similares.
- Experimento 6. Detección de objetos en el signo: en ocasiones, la imagen de un signo puede venir acompañada de un objeto, como por ejemplo la presencia de uno a más anillos en los dedos o un reloj o una pulsera en la muñeca, o porque se sostiene un objeto como puede ser un bolígrafo. Se ha medido la tasa de éxito teniendo en cuenta esta situación.
- Experimento 7. Rendimiento del algoritmo: el rendimiento o exactitud del sistema también se evalúa mediante la obtención de una matriz de confusión. Como se ha mencionado en un epígrafe anterior, se alcanzan cifras de acierto en torno al 90%, incluso 95% si el usuario se ha entrenadoen el uso del sistema.

- Experimento 8. Creación de una nueva base de datos: se ha intentado explotar la similitud entre signos en distintas lenguas de signos al objeto de crear una base de datos única para varias lenguas. Se ha encontrado, obviamente, que el sistema entonces no es capaz de identificar lalengua de signos que está siendo usada.

Diferencias con otros sistemas

En este trabajo de investigación, la capacidad otros sistemas anteriores se mejora significativamente y las nuevas características incluyen bases de datos conjuntas de lenguas de signos combinadas, ASL, PSL y SSL. Las versiones de estas bases de datos se mejoran al incluir más imágenes de mejor calidad y variaciones de gestos individuales. El sistema actualizado ahora incluye una función de detección de idioma y un perfil de usuario mejorado que contiene más información como la mano derecha/izquierda, la velocidad y exactitud de los gestos de la mano del usuario, la base de datos de lenguaje de signos del usuario por defecto, etc. El sistema propuesto se valida aún más utilizando nuevos experimentos. También se realiza una comparación de los tres lenguajes de signos. Se amplía una revisión de la literatura para ayudar a mejorar el sistema propuesto. La aplicación de software se actualiza y incluye las características mencionadas aquí. Algunas nuevas características, es decir, detectar gestos cuando el usuario está sosteniendo o usando otros objetos como un reloj de pulso, anillo, etc., también se añaden. La base de datos se actualiza con nuevas imágenes, y el algoritmo se mejora aún más para aprovechar las nuevas características.

Muchos sistemas similares existentes sólo se centran en las personas D-M, donde se les proporciona la interfaz para que se detecten sus gestos de mano. Estos sistemas esperan que las personas de ND-M aprendan sobre la identificación de signos y las interfaces que se les proporcionan. No es fácil para las personas de ND-M aprender lengua de signos, reduciendo así la eficacia de este sistema. El sistema propuesto ofrece una solución completa a los problemas que enfrentan tanto las personas de D-M como de ND-M.

Trabajos futuros

Como se describe en la Memoria, posibles trabajos futuros son:

- El aumento de los signos contenidos en las bases de datos de las distintas lenguas de signos.
- El aumento de las lenguas de signos que puede manejar el sistema, aprovechando las similitudes para aumentar su eficacia.
- Mejora de los algoritmos de reconocimiento de signos.
- Mejora de los mecanismos de entrenamiento.

Desde el punto de vista académico, el tercer punto es el más prometedor, siendo los otros tres más orientados a la ingeniería y a mejorar una herramienta que se ha demostrado de inmensa utilidad para facilitar la comunicación de personas con discapacidad auditiva y, por consiguiente, en riesgo de exclusión social.

CONTRIBUCIONES DE ESTA TESIS DOCTORAL

Las principales contribuciones de esta tesis doctoral se encuentran descritas en los capítulos 3 y 4 de esta Memoria. Se exponen brevemente en este epígrafe, como conclusión de este resumen.

El capítulo 3 contiene lo adelantado en el artículo:

Saleem, M.I.; Siddiqui, A.; Noor, S.; Luque-Nieto, M.-A.; Otero, P. A Novel Machine Learning Based Two-Way Communication System for Deaf and Mute. Appl. Sci. 2023, 13, 453. https://doi.org/10.3390/app13010453.

En el artículo se describe el sistema del que ya se ha hablado en este resumen y se explica que hasta la fecha no se habían utilizado las técnicas de aprendizajes profundo y automático supervisado (*deep learning* y *supervised machine learning*) para la adquisición de signos de la lengua de signos. Se ha descrito cómo se aplican estas técnicas a ese problema, se evalúa experimentalmente su tasa de éxito y se concluye que se trata de dos técnicas que conducen a muy buenos resultados. Parte del éxito de los experimentos se deben a la idea propuesta de que el sistema aprenda de las propias personas D-M que lo utilizan, es decir, de signos realizados precisamente por los usuarios del sistema.

De igual manera, en el capítulo 4 se desarrollan los contenidos del artículo:

Saleem, M.I.; Siddiqui, A.; Noor, S.; Luque-Nieto, M.-A.; Nava-Baro, E. A Machine Learning Based Full Duplex System Supporting Multiple Sign Languages for the Deaf and Mute. Appl. Sci. 2023, 13, 3114. https://doi.org/10.3390/app13053114 (JCR-Q2).

Esta contribución consiste principalmente en cómo aplicar las técnicas desarrolladas en la contribución anterior para que el sistema sea capaz de soportar diferentes lenguas de signos y también se ha usado un algoritmo de red neuronal convolucional (*convolutional neural network*). Para ello, como ya se ha comentado, se crearon tres bases de datos de imágenes de signos con un total de imágenes de alrededor de 6000 y se han explotado las similitudes existentes entre las distintas lenguasde signos. Además, se han tratado dos importantes características: la conversión automática de signosa voz y su inversa y la posibilidad de entrenar a los usuarios D-M para que mejoren la calidad de sus signos con el objeto de elevar la tasa de éxito. El reconocimiento de voz se ha completado con la creación visual de signos. Al igual que en el capítulo 3, se ha evaluado experimentalmente la tasa de éxito que, en algunos usuarios, ha llegado al 95%.

COOPERACIÓN HISPANO-PAKISTANÍ

Un asunto no menor en absoluto, es que los trabajos de investigación conducentes a la realización y finalización de esta tesis doctoral se han llevado a cabo en el marco -no oficial- de una cooperación entre la Universidad de Málaga y las universidades pakistaníes *Sir Syed University of Engineering & Technology* (SSUET), Karachi, Pakistán, y *Mehran University of Engineering & Technology* (MUET), Jamshoro, Pakistán. En concreto, el autor de la presente tesis es miembro facultativo de la primera de ellas.

Esta cooperación, llevada a cabo hasta la fecha, casi sin financiación pública, se inició como fruto de la actividad del malogrado profesor titular de la Universidad de Málaga, Dr. Javier Poncela, a raíz de su nombramiento como Subdirector de Relaciones Internacionales de la ETSIT-UMA. Sólo una beca predoctoral de la *Higher Education Commission* (HEC) de Pakistán ha sido la contribución económica de ambos países, que ha dado como fruto una tesis doctoral con mención internacional defendida en 2023 en la UMA.

Ha sido la financiación europea, a través de sus programas KA107 y KA2 la que ha permitido que el excepcional desempeño del Dr. Poncela haya culminado en la finalización hasta la fecha de siete tesis doctorales (la presente será la octava) en un plazo de seis años. Cifra asombrosa si se tiene en cuenta las dificultades a las que se han enfrentado en estos años el mundo en general (crisis económica, pandemia, guerras, dramas de desplazados, …) y los investigadores en particular (alumnos de doctorado, directores de tesis y tutores académicos). Y todo ello sin menoscabo de la calidad de las contribuciones, calidad cuyo umbral está fijado por la Agencia Española de Investigación.

Por ello, todos aquellos que han contribuido a estos resultados desean dejar constancia de la deuda que con el Dr. Javier Poncela tienen las instituciones en que sirvió con lealtad y generosidad. Y confían en que en un futuro próximo esa dedicación se vea premiada con nuevos frutos y con una más intensa cooperación también a nivel institucional entre ambos países.

This page intentionally left blank

Acknowledgements

First and foremost, I would like to praise Allah for granting me the opportunity and fortitude to undertake this difficult endeavor. This Ph.D. dissertation was completed in 2023. I would like to thank everyone at the University of Málaga for allowing me to carry out my research work and complete this doctoral dissertation.

I would like to salute my mentor, Prof. Dr. Pablo Otero Roth and my supervisor Prof. Dr. Miguel- Ángel Luque-Nieto for their unwavering support and motivation. I would also like to mention Prof.Dr Javier Poncela who left us few years ago. I would not have completed this work without his contribution. My supervisors gave me sufficient time to begin my computer vision research and had faith that I would accomplish something. They gave me as much latitude as I required to complete my research.

My sincere appreciation also extends to Sir Syed University of Engineering and Technology, particularly to my teachers, Dr. Atif Siddiqui and Professor Dr. Muhammad Aamir, for their encouragement and support. You inspired me to begin this new chapter in my life. I couldn't have done it without you. I will always remember attending your classes. You are the best teachers. I want to thank Dr. Atif Siddiqui in particular because this accomplishment would not have been possible without his constant help. In this world, only a few people will be there for you when you need them the most.

Last but not least, I would like to thank to my family. My father, Squadron Leader (Rtd) Muhammad Saleem Ansari (late) who served Pakistan Air force and was a professor of English insisted that I should get my doctoral degree (May God bless him and may you pray for him). My mother Mrs. BanoSultana, who stood by me in all adversity, is a very caring mother. My sisters, Dr Saima Saleem, Dr Ayesha Nadeem and brother in law Mr Nadeem Iftikhar, my wife Dr Shaheena Noor all supportedand made this possible. Last but not the least I would like to mention my nephews Muhammad Umer and Muhammad Usman who always encouraged me with their cute, little smiles.

This page intentionally left blank

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ASL	American Sign Language
ArSL	Arabic Sign Language
ArgSL	Argentinian Sign Language
AI	Artificial intelligence
ANN	Artificial neural networks
BSL	Bangla Sign Language
CSL	Chinese Sign Language
CPI	Chronological-Pattern Indexing
COTS	Commercial-off-the-shelf
CNN	Convolutional neural networks
D-M	Deaf and mute
DBN	Deep belief networks
DTW	Dynamic Time Warping
FSL	Filipino Sign Language
GMM	Gaussian Mixture Model
GAN	Generative adversarial networks
GUI	Graphical user interface
HMM	Hidden Markov Model
ISL	Indian Sign language
ICT	Information and communication technology
KNN	K-Nearest Neighbours
LMD	Leap Motion Device
LMC	Leap-Motion-Controller
LSTM	Long Short-Term Memory
ML	Machine learning
MLP	Multilayer perception
NLP	Natural language processing
ND-M	Non-deaf and mute
PSL	Pakistani Sign Language
PCA	Principal Component Analysis
RNN	Recurrent neural networks
RGB	Red, Green and Blue
RSL	Russian Sign Language
SLD	Sign Language Dataset
SLI	Sign Language Interpreter
SLR	Sign Language Recognition
SSL	Spanish Sign Language
TSL	Turkish Sign Language
VRS	Video Relay Service
WHO	World Health Organization

Chapter 1: Introduction

1. Introduction

Technology has become an integral part of our lives. This has a positive impact on society, and people benefit from it. This research is focused on helping people with impairments. Artificial intelligence can support people with impairments to integrate into society effectively. The World Health Organization (WHO) reports that 5% of the world's population is deaf and mute (D-M) [1].

There are millions of D-M people who heavily rely on sign language to communicate with others. They are often faced with the challenge of establishing effective communication when others are unable to understand or use sign language. In general, people with no physical impairment do not make an effort to learn sign language. This creates a scenario where it is difficult to establish effective communication between the D-M and non-deaf and mute (ND-M). To overcome this problem, a system is needed that provides a platform where the D-M and the ND-M can communicate with each other.

This work focuses on the use of information and communication technology (ICT) to provide a platform for which a system can be developed to support this type of communication. The research covers the development of a software application that is interfaced with hardware and uses different algorithms to solve some of the above-discussed problems. This is expected to be a step towards improving the integration of the D-M in society. The development presented in this research work uses existing sign language datasets to establish automated communication. The proposed system is designed around commercial-off-the-shelf (COTS) hardware integrated with a software application using machine learning (ML) algorithms that provide full-duplex communication.

The use of ML is providing a growing number of solutions to complex problems in diverse areas related to human impairment. ML is already applied to establish effective communication systems so that the D-M and the ND-M can communicate. Researchers are using existing algorithms, as well as creating new algorithms, to enhance the existing systems.ML algorithms are not only being deployed to solve existing problems, but these algorithms are also employed to increase efficiency by reducing processing time, creating consistent datasets, and improving quality.

1.1. Sign Language

The method of communication between D-M and others is through sign language, which can only be effective when individuals from both sides understand sign language. This research is a step toward presenting a solution that will improve this communication. One of the main problems highlighted is the lack of understanding of sign language by ND-M people.

Alternate solutions are considered where one is to use writing for communication. This is an alternative method of communication, but it is thought to be a sluggish and inefficient method [1]. Another solution is to use a sign language interpreter, which is realistic yet expensive, and availability is a bottleneck. This option is also not available to the majority of the D-M community. Due to these and similar issues, D-M people find it hard to communicate with others and are thus unable to integrate into society. This means that society is unable to benefit from its contribution.

Having a smart model that allows communication between D-M and ND-M people will provide a breakthrough in tackling this problem and will have a huge impact on society due tothe integration of more people and their contributions. The research work presented here takesinto consideration the work carried out in [2]. A two-way smart system for communicating between D-M and ND-M people is presented. There are two major contributions in this work, which are (a) Leap Motion for gesture recognition, and (b) an Android application. The Leap Motion device is used to recognize hand gestures, which are in use in Pakistani Sign Language (PSL) courses.

1.2. Artificial Intelligence

The use of artificial intelligence (AI) is increasing, and researchers are solving various problems through its application. This research work also focuses on AI and how it can be used to provide solutions to D-M people.

1.2.1. Machine Learning

In this research, supervised machine learning is applied, and a dataset is created to provide a new feature for D-M people so that they can communicate in more than one language. Some machine learning algorithms are support vector, linear and logistic regression, and support vector.

1.2.2. Deep Learning

The detection of hand gestures and processing of these data are the most important options required for any system that supports D-M. For this research, artificial neural networks(ANN) and convolutional neural networks (CNN) are reviewed.

1.2.3. Artificial Neural Network

ANN is used to tackle a wide variety of computer vision problems, including classification, in today's world. Researchers who are tasked with collecting, analyzing, and interpreting vast amounts of data will find that ANN are very beneficial to their work. In addition, ANN increases processing speed and reduces complicity. The application of AI can be seen in areas such as image processing, natural language processing, intelligent robots, knowledge representation, and automatic reasoning. Some examples of algorithms based on neural networks are multilayer perception (MLP) and Boltzmann neural networks.

ANN is a learning system that models human brains. It does it by deploying a networked system of nodes or neurons that are arranged in layers. Learning from previous experiences allows the system to recognize patterns, organize data into categories, and forecast future occurrences.

A fundamental ANN model is presented in Figure 1.1. The structure has one input payer which has an input feature vector of length 'n'. At the output there is also one layer with output variables of length 'm'. The input features are processed through to the output layer viahidden layer of neurons of length 'k'. The raw data i.e., the data to be processed is entered through input layer. This layer only provide interface for the data which is required to be processed. The processing is then carried out by the hidden layer. The hidden layer is the key to this network where all computational processing of the characteristics acquired from input layer is done. The processed data is then forwarded to the end user via output layer. The data sent to the output layer is the solution to the problem interfaced from input layer.

The behavior of the system is determined by the way the nodes are coupled, and the weights are modified. The network first goes through training. During the training process, these weights are automatically adjusted using a learning algorithm to ensure that the ANN performs the task in an optimal manner.

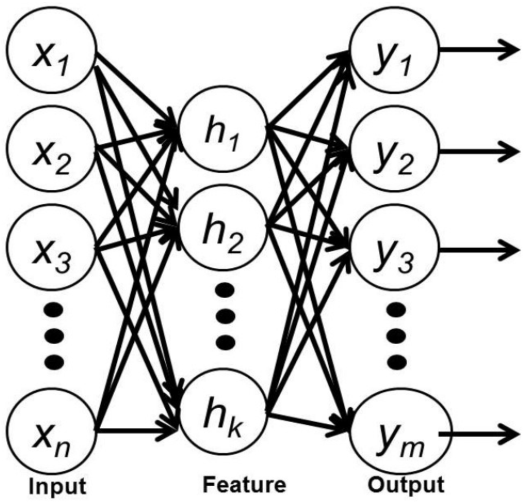


Figure 1.1. General architecture of neural network.

An important parameter of ANN is an activation function. This provides necessary processing required for the ANN to learn a complex patterns within the input data. The process replicates how a human brain works. Following are some commonly used activation functions [3].

The Binary Step Function: The activation function of a neuron is governed by a threshold value, which determines whether the neuron should be activated. If the value is greater than the threshold, the neuron is activated; otherwise, the neuron remains deactivated, and itsoutput is not passed on to the next hidden layer. The threshold is the main feature of this function which can be set to a value depending on the requirement.

The linear activation function: The activation function is proportionate to the input. This function is like an analog system.

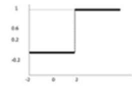
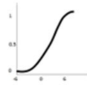
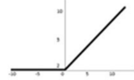
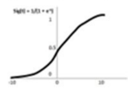
The Sigmoid or Logistic Activation Function: This function takes a real number as an input and returns a zero or one at the output. The output is a one if the input is high i.e., close to maximum while the output is a zero if the input is close to the minimum. The maximum and minimum values for the input are defined through the range.

The Tanh Function: This function is also known as the Hyperbolic Tangent (HT). This is like the sigmoid and logistic activation functions, with the exception that the output range for the Tanh function is negative one to one rather than zero to one. When the value of the input is close to maximum then the output is one while the output is negative one if the input is close to minimum limit.

Rectified Linear Unit Function (ReLU): This function is a derivate function even though the name suggests this is a linear function. This function is computationally efficient and allow back propagation. The main limitation of this function is that it does not excite all of the neurons at the same time.

The Softmax Function: The softmax function takes a vector of 'k' real values at the input and offsets these values with aone. The softmax algorithm takes different input values like positive, negative, zero, or even more than one and convert into zero or one. Due to this feature, it is possible for the data to be interpreted in terms of probability. Table 1.2.3.1 presents a summary of different activation functions discussed in this section

Table 1.1 Summary of activation functions including mathematical presentation. [3]

Activation Function	Mathematical Formulation	Diagram
BinaryStepFunction	$f(x) = \begin{cases} 0, & \text{if } x < 0. \\ 1, & \text{otherwise} \end{cases}$	
LinearActivationFunction	$f(x) = x$	
Sigmoid/LogisticActivationFunction	$f(x) = \frac{1}{1 + e^{-x}}$	
TanhFunction	$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	
ReLUFunction	$f(x) = max(0, x)$	
SoftmaxFunction	$softmax\ (z_i) = \frac{exp(z_i)}{\sum_j exp(z_j)}$	

1.2.4. Convolutional Neural Network

Other deep learning-based algorithms include CNN, recurrent neural networks (RNN*)* [3], generative adversarial networks (GAN), and deep belief networks (DBN) [3]. One of the important aspects of this research is to process image data, i.e., hand gestures from D-M. For processing image data, CNN is used [3].

There are number of architectures that are created and developed using CNN, e.g., LeNet, AlexNet, Google Net etc. These algorithms are limited in their applicability in portable devices due to their memory and processing requirements. A CNN requires high processing time and GPU. To handle this high computational data, small CNNs are created which are called Mobile Nets that uses convolution on the smaller CNNs to increase classification accuracy without increasing network parameters. The key feature is that this is compatible with mobile or embedded devices.

A Mobile Net streamlined architecture that comprises of lightweight CNNs using depth wise separable convolutions. Figure 1.2 shows the architecture of Mobile Net focusing on depth wise separable filters.

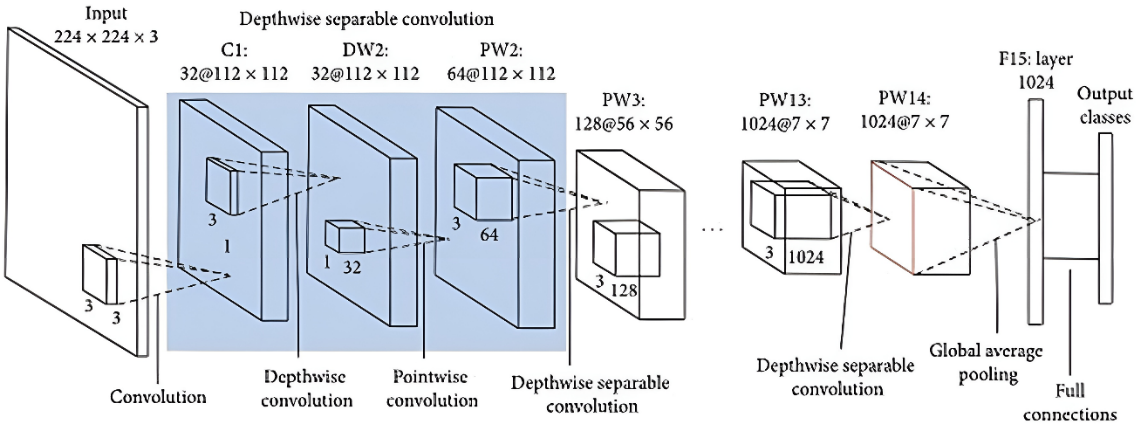


Figure 1.2. Mobile Net Architecture.

Depth-wise separable convolutions are the building blocks of the Mobile Net architecture. These are created by combining the depth-wise and point wise convolution filters. Mobile Nets apply depth-wise convolution as a single filter to each input channel. The results are then mixed with the results of the point wise convolution using a 1x1 convolution scheme. The processing time is reduced by filtering and organizing the data by using a 2-layer factorization. The architecture is presented in figure 1.3.

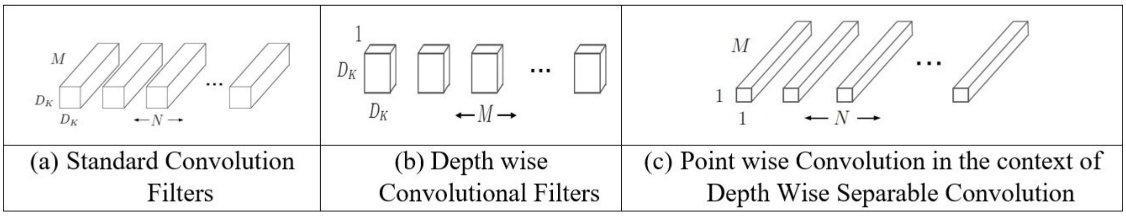


Figure 1.3. The standard convolutional filters.

1.2.5. Scope of Research

This section presents the scope of this research work. Figure 1.4 shows the scope of this research within the AI domain. The two solutions are implemented using supervised machine learning and CNN.

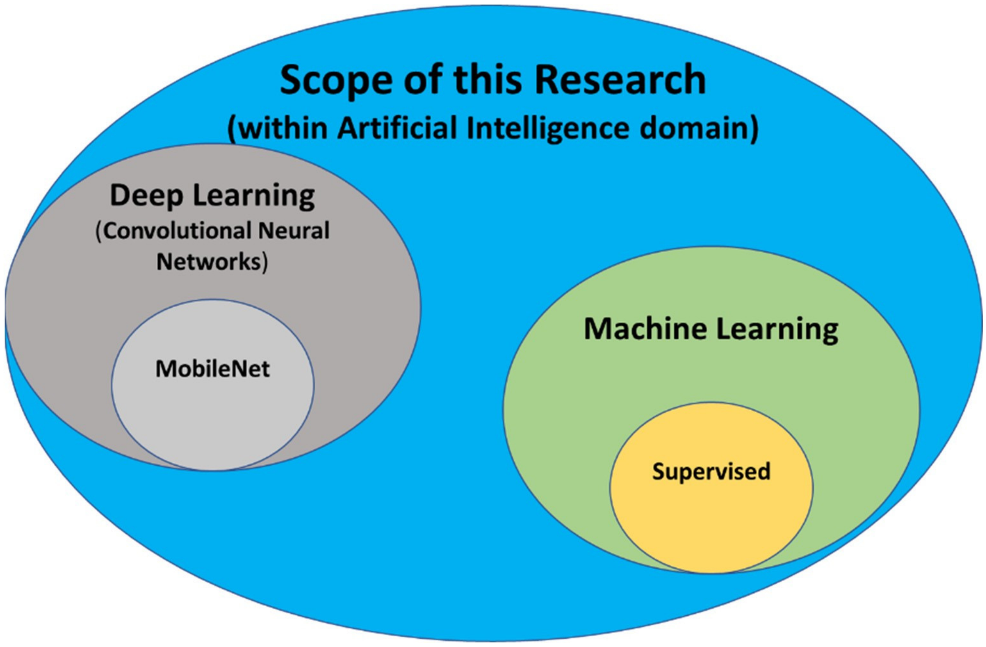


Figure 1.4. Scope of this research within the artificial intelligence domain.

In this research work, the capability of the previous system [4] is significantly enhanced, and new features include the new combined sign language datasets, American Sign Language (ASL), PSL and Spanish Sign Language (SSL). The online versions of these datasets are enhanced by including more images of better quality and variations of individual gestures. The updated system now includes a language detection feature and an improved user profile which contains more information such as right/left hand, speed and accuracy of user's hand gestures, default user's sign language dataset, etc. The proposed system is further validated using new experiments. A comparison of the three-sign languages is also carried out. A literature review is extended to help improve the proposed system. The software application is updated and includes the features mentioned here. Some new features, i.e., detecting gestures when the user is holding or wearing other objects like a wristwatch, ring, etc., are also added. The database is updated with new images, and the algorithm is further improved to exploit the new features.

In this research work, limitations of the existing systems are studied, and a new two-way communication system for the D-M and the ND-M is presented based on a new dataset created, which is the combination of ASL, PSL, and SSL, with a focus on overcoming limitations in existing systems. Within this research, the three datasets which are individually available are further enhanced to improve the quality of these datasets. The details of the proposed system are presented in this thesis and validated using a series of experiments. The acquired hand gesture data is processed using CNN, and further processing, discussed later, is implemented using supervised ML.

Figure 1.5 shows the key areas of this work, which encompasses the four main building blocks, which include CNN, supervised ML algorithms, selection and processing of sign language datasets, and review of hardware tools for hand gesture acquisition.

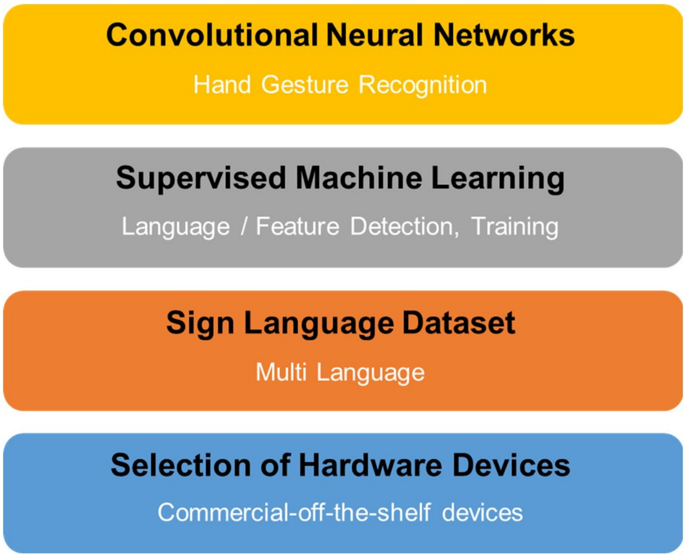


Figure 1.5. Key areas of this work.

Here in above (fig 1.5.key areas of this work) the convolutional neural networks (CNNs) serve as a highly potent algorithm utilized within the realm of network architecture to facilitate the process of image classification. By harnessing a multitude of diverse characteristics, CNNs enable the effective differentiation among various distinct classes, thereby enhancing the overall efficacy of the classification task. Supervised learning (SL) trains a model utilizing numerous pairings of an input item (e.g., a vector) and a desired output value (also known as human-labeled supervisory signal).

Chapter 2: Related Work

2. Literature Review

This section is a review of recent research work carried out for the development of communication systems involving the D-M. The review includes the novel features presented by the authors as well as the limitations of their proposed systems.

Figure 2.1 shows the review carried out within the categories. The researchers used one or more of these categories for implementing their systems.

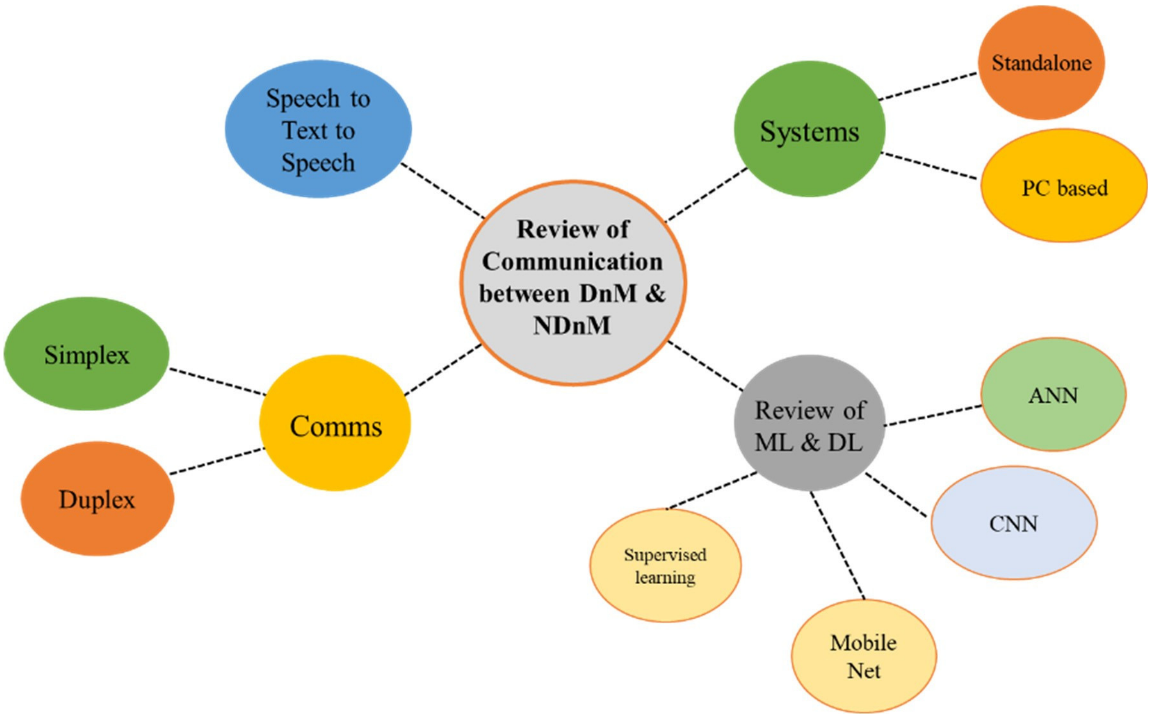


Figure 2.1. Literature review map.

2.1. Overall Literature Review

A fundamental ANN model is presented in [5]. The structure has one input payer that has an input feature vector of length ‘n’. At the output, there is also one layer with output variables of length ‘m’. The input features are processed through to the output layer via a hidden layer of neurons of length ‘k’. The raw data, i.e., the data to be processed, is entered through the input layer. This layer only provides an interface for the data that is required to be processed. The processing was then carried out using the hidden layer. The hidden layer is the key to this network, where all computational processing of the characteristics acquired from the input layer is done. The processed data are then forwarded to the end user via the output layer. The data sent to the output layer are the solution to the problem interfaced from the input layer. An application of the Mobile Net is presented in [6]. This is a streamlined architecture that comprises lightweight CNNs that use depth-wise separable convolutions.

The first manuscript [7] presents different recognition systems for sign languages. The authors in [8] presented a communication framework for D-M people. They used an automated Sign Language Interpreter (SLI), where they acquired data using motion sensors fitted to a glove. The sensor data are acquired using an Arduino board. The acquired data are then processed using a machine-learning technique. The system achieves an accuracy rate of 93%. In [9], a speech recognition system is presented that identifies a registered speaker’s voice. This is a computer-based implementation in which language patterns are stored. The information was processed and acquired using a questionnaire created by the authors in [10]. They collected data from various organizations set up for D-M people. They concluded that the average age when hearing loss is discovered is 2.8 years, while a hearing aid is normally used from 7.6 years of age.

A system is presented in [11], which is referred to as AWAAZ. D-M people can use this system, where their gestures are acquired using image processing. The main components ofthe gesture recognition system are image acquisition, segmentation, morphological erosion, and feature extraction. In [12], a system that supports two-way communication between D-M and ND-M people is presented that uses an automatic speech recognition technique and a mobile application based on visualization. The authors developed an Android-based application in [13]. This uses the Principal Component Analysis (PCA) approach and the Video Relay Service (VRS). The VRS works as a manual interpreter, converting speech to hand signals and vice versa. The hand gesture is recognized by the PCA algorithm.

A web cam-based application is developed in [14], where the images are first acquired using a webcam, then PCA is applied to extract features and the characters are recognized using training sets. The authors surveyed [15] and published the issues faced by D-M children. Theyconcluded that it is possible to apply a common solution due to various parameters. Some solutions are presented in this paper. They also concluded that Hand-Gesture-Recognition (HGR) systems have contributed significantly to a shift in the way people interact with computers.

There is a lot of focus on improving sensor technology, which is expected to have a significant impact on how D-M people can communicate. The Leap-Motion-Controller (LMC) [16] is one such device. Despite the success of developing cutting-edge algorithms, there are still limitations because these algorithms have not focused on how to quickly processsequential hand gesture data and are unable to characterize the discriminative representation of different classes of hand gestures. The LMC sensor acquired hand motions and hand motion data where the pattern is sorted using a novel Chronological-Pattern Indexing (CPI) method. This approach interprets the data as a series. Hand gesture recognition systems have recently been demonstrated to offer substantial promise for use in the realm of digital entertainment. These enhancements have been made possible because of recent advancementsin machine learning and sensor upgrades. Contactless hand gestures on a Leap Motion device are used to identify dynamic hand motions. This methodology is within the scope of this research study; hence, it has been included and reviewed in detail.

In [17], the authors used recurrent neural networks in combination with Long Short-Term Memory (LSTM) to analyze sequential time series data acquired through the Leap Motion device for gesture recognition. The authors used both the normal unidirectional LSTM and the bidirectional LSTM. A prediction network architecture, identified as the Hybrid Bidirectional Unidirectional LSTM, is created based on the model discussed here, combined with other components. This model significantly improved performance in both forward and reverse directions by considering the spatial and temporal interactions between the Leap Motion data and the network layers.

A portable D-M sign language translator is presented in [18]. This uses a controller to analyze the gestures in a photograph using various image-processing techniques. The use of deep learning has greatly enhanced the performance of this system. After completing the detection process, the hand gesture signs are converted into spoken language. There is a limitation to this application, which is also highlighted. The cost of converting this into a commercial product is high. Researchers are currently placing greater emphasis on the development of Sign Language Recognition systems that are suitable for commercial use [19]. There are several methodologies applied to find a solution to reduce costs. The main building block is data acquisition. The cost of data acquisition devices is generally high. There is a focus on finding a solution to reduce this cost, which will reduce the overall cost of a commercial product. Another aspect of the research is to review and evaluate how Sign Language Recognition (SLR) can be improved. Different SLR systems being developed have their pros and cons. In [20], the authors developed a glove using flex sensors. The hand gesture data from the glove are acquired using an Arduino board.

A leap motion controller is an interactive tool with many applications, such as manufacturing, 3D modeling, and other fields. This is also true in detecting hand gestures. It follows the movement of the hand with the help of cameras and infrared LEDs [21]. The device is used in combination with the SDK development toolkit, which collects hand gesture data. The results are favorable. The focus of the research in [22] is acoustic communication. The authors highlighted the difficulties that D-M people face in communicating with others. A real-time device that aids D-M people, where they use a specifically designed glove fitted with five flex sensors and one accelerometer. The sensor output will change based on hand gestures. The acquired data are then processed using an Android application. This application can interpret gestures in Arabic Sign Language (ArSL). The output is in the form of both text and voice. The prototype is a reliable and accurate device developed at a low cost.

The importance of sign language is highlighted in [23]. The focus is also on ND-M people, as they also need to learn sign language so that they can communicate with D-M people. The available material is less effective in terms of training sign language for ND-M people. The authors developed an interactive Chinese sign language teaching system for smartphones that makes it easier to learn sign language. The application teaches sign language in an interactive way, where a mobile phone camera is used to record the activity. The researchers also added a vocabulary of 100 words. The application also carries out analysis and assessment, which helps the user.

In [24], a mobile phone application 'BridgeApp' is developed for D-M people. They used speech-to-text, speech-to-visual, and sign language features in their implementation. This application works in offline mode. They have implemented ASL and Filipino Sign Language (FSL). A system based on vibrotactile, and visual feedback is developed in [25]. D-M people can communicate on their own without the use of sign language using this system. An Android application is also developed by the authors.

The authors in [26] used ARM LPC-2148 to connect different sensors and actuators to a Braille keypad, which is a user-friendly application for blind people. They have also installed sensors in walking sticks to aid people who are visually impaired. A Hand Gesture Recognition System for D-M people is designed using a Fuzzy-Neural Network [27]. Thedata was acquired from a hand gesture translating glove. The accuracy of gesture recognition using fuzzy logic is discussed. In [28], an application is developed for D-M, which detects facial emotion. The application performs eye-tracking and event-related potential analysis. The authors carried out extensive testing using 630 images of hand gestures.

A deep learning-based hand gesture translation system is presented in [29]. The software automatically recognized hand gestures at 94% accuracy. A process to convert to and from sign language is discussed in [30]. The gestures are detected and then converted into text. To develop an effective model, a variety of ML and AI techniques are used, in addition to naturallanguage processing (NLP) and convolutional neural networks CNN). The authors also used attention-based LSTM to detect rapid and continuous lip movements.

An intelligent glove for sign language communication is discussed in [31]. The researchers used a flex sensor and a GY-521 module interfaced with Arduino. The flex sensor measures the movement of the fingers and generates a voltage output. The acquired hand gesture data are then converted into text. In [32], the authors developed a module that translated sign language. The user can wear it around the neck. The device processes hand gesture images using image processing techniques and deep learning models to recognize the sign, which is then converted into a voice using a text-to-speech convertor.

A CNN-based system is presented in [33], which also provides a training option. This system also supports multiple languages. In [34], a low-cost Arduino-based system connected to flex sensors is presented. This system is used for gesture detection. The system also uses agyroscope and an accelerometer. A picture-based communication application is presented in [35]. It is a portable, easy-to-use communication tool that helps people communicate witheach other. A Chinese language translation and processing tool is presented in [36]. The D-M people can watch videos with Chinese subtitles using this tool.

In [37], a video chat application is developed. This is based on the Indian Sign language (ISL). The process is initiated when a user starts making hand gestures, which are then picked up by a camera. The algorithm, which is based on CNN, then converts this into phrases and numbers. A medical consultation system is presented in [38], where the D-M person can attend a hospital and communicate with the doctor. The system provides two-way communication. In [39, 40], the authors applied machine learning to solve a new problem for setting up electronic product test sites. They created a dataset for the machine learning algorithm.

The creation of a sign language dataset is outside the scope of this work. The following references provide the details of some available sign language datasets.

In [41], the authors presented the Argentinian sign language dataset. They included 3200 videos in the dataset. They highlighted the importance of having the entire dataset for training purposes. The authors in [42] proposed a large-scale ASL dataset that includes 25,000 videos. A Chinese sign language dataset is collected and processed by the authors in [43]. The dataset includes 500 categories and was evaluated by the authors, including another dataset. The authors in [44] created a new Russian sign language dataset. A new Turkish sign language dataset is presented, which includes 226 signs that the authors processed using CNN [45]. In [46], the authors presented a word-level ASL dataset. The authors in [47] worked on word-level ASL and collected references related to dataset creation. In [22], the authors focused on Arabic sign language and created a glove for detecting this dataset.

Table 2.1.1 is a summary of the literature review and details of the hardware platform and software, or algorithm used by the researchers. The last column highlights some features of the work carried out.

Table 2.1.1: Summary of Overall Literature Review.

References	Hardware	Software/Algorithm	Features/Limitations
[8]	Gloves, Sensors, Arduino	KNN	Custom made, require validation; D-M to ND-M only
[9]	PC based	None	ND-M to D-M to only
[10]	None	None	Survey only
[11]	PC based	Image processing	D-M to ND-M only
[12]	Mobile app	Speech recognition	D-M to ND-M only
[13]	Mobile app	ANN	D-M to ND-M only
[14]	PC based	PCA	D-M to ND-M only
[15]	None	None	Survey only
[16]	Leap Motion	CPI	D-M to ND-M only
[17]	Leap Motion	LSTM	D-M to ND-M only
[18]	Raspberry Pi	Deep learning	D-M to ND-M only
[19]	PC based	Various	Survey only
[20]	Gloves, Sensors, Arduino	None	D-M to ND-M only
[21]	Leap Motion	None	D-M to ND-M only
[22]	Gloves, Sensors, Mobile app	None	D-M to ND-M only
[23]	Mobile app	None	D-M to ND-M only
[24]	Mobile app	None	D-M to ND-M only
[25]	Mobile app	None	D-M to ND-M to D-M
[26]	Sensors, Standalone	None	D-M to ND-M only
[27]	PC, Microcontroller	FNN	D-M to ND-M only
[28]	Gloves, Microcontroller	None	D-M to ND-M only
[22]	PC based	Deep learning	D-M to ND-M only
[23]	PC based	CNN	D-M to ND-M only
[24]	Gloves, Sensors, Arduino	KNN	D-M to ND-M only
[25]	PC based	NN	D-M to ND-M only
[26]	PC based	CNN	D-M to ND-M only
[38]	Mobile app	Mobile Net	D-M to ND-M to D-M

2.2. System Level Review

In the manuscript [48], the authors used a Leap Motion Device (LMD) to acquire hand gesture signals. The D-M can communicate with others using hand gestures while the ND-M are provided with an android application that converts speech to text. The accuracy of gesture detection is 95%. The authors in [11] presented a system that converts hand gestures from the D-M into text. The system also displays a gesture image for the text input. A portable device is designed by the authors in [18]. The D-M can carry this device to acquire and convert hand gestures into speech. In [17], a contactless hand gesture recognition system is presented where the gesture is acquired using an LMD. LSTM recurrent neural networks were used for gesture detection.

The authors used an LMD for hand gesture recognition for the D-M in [16]. They defined a criterion for hand gesture recognition and used different methods such as Gaussian Mixture Model (GMM), Hidden Markov Model (HMM), and Dynamic Time Warping (DTW) for image data processing. In the manuscript [19], the authors acquired sign language gestures using a camera. Various neural network algorithms were compared to process the image data and calculate the accuracy of the proposed system. The authors in [5] developed a learning system for the D-M people. They used Augmented Reality to convert the acquired gesture into a 3D model. They also developed a glove using Flex sensors.

The manuscript [10] is a survey paper where the authors collected and reviewed data to understand different reasons for hearing loss and the age when it is detected. A prototype of a glove using flex sensors is designed by the authors in [20]. They designed a system where hand gestures are acquired and then converted into text and voice. The researchers in [49] reviewed and compared several techniques for hand gesture recognition and listed the novel features of different algorithms, and they also highlighted their limitations. This surveyprovides a good insight into different algorithms, including their pros and cons. In [50], an automated ASL-based hand gesture recognition system is presented, where hand gestures are first detected and then converted into text. An automated sign language detection system is proposed by the authors in [51], where they designed two-way communication using an ASL dataset.

The Bangla Sign Language (BSL) dataset is applied in [52], where the authors processed the hand gesture data using CNN, reporting an accuracy exceeding 98%. In [36], the authors developed a mobile phone application where a customized interface for individual users is proposed. This application provides support for multiple languages. An ArSL based system is presented by authors in [53]. The detected ArSL hand gestures are converted into voice. In this manuscript, the authors surveyed different hand gesture detection techniques.

A CNN-based hand gesture recognition system is presented in [29]. The authors reported a hand gesture recognition accuracy exceeding 93%. In [54], the researchers developed a glove that was used for translating ASL and reported an accuracy of 95%. A portable hand gesture recognition prototype is presented in [32]. The authors of this manuscript used a deeplearning algorithm to process image data, which is then converted to speech. In [55], the authors developed a prototype of a smart glove for detecting sign language, and when a gesture is recognized, a recorded audio message is played. This system is designed to detect ISL data. Another CNN-based hand gesture detection prototype is presented in [56]. The authors claim that the proposed prototype detects ISL with training accuracy exceeding 99% by using more than 3000 images.

In [7], the authors used the ASL dataset, where the hand gestures are detected using CNN. They reported an accuracy exceeding 95%. The authors in [31] developed a prototype using aflex sensors-based glove and interfaced it with Arduino. The prototype converts the hand gesture data into text. The authors used K-Nearest Neighbors (KNN) algorithm for imagedata processing. A PC-based image detection system is presented in [14]. The authors usedthe PCA algorithm. In [8], an automated SLI prototype, which is based on a custom-made glove connected to Arduino, is presented. The authors reported a hand gesture detection accuracy of 93%. A CNN-based prototype for hand gesture detection is presented in [33].This prototype also provides a training option and support for multiple languages.

The authors developed a mobile application in [24] for the D-M. The application works in offline mode and uses ASL and FPL datasets. A PC-based prototype is presented by authors in [34] for detecting hand images using sensors. An android-based application is presented in [25], where visual feedback is provided to the user. The authors in this paper also compared different technologies. In [15], the authors carried out a survey to understand the problems and challenges faced by the D-M. They highlighted the importance and impact of hand gesture recognition systems and prototypes in providing a platform for the D-M.

In [57] the authors presented a 3D hand gesture detection system using LMD. They used fingertips and palm position to detect hand gesture and processed through SVM. Theyreported 81% accuracy. They also created a dataset of 550 samples. The authors in [58] presented a use for LMD device. The authors used Java for developing their application and converted image acquired data using machine learning model. In [59] another application is developed using LMD. The authors integrated LMD, sensors and Arduino board to remotely control a robot which is used for humanitarian purposes.

A numeral gesture recognition system is presented based on LMD in [60]. The authors published an accuracy of 70%. The detected gestures are drawn in free air within the range of LMD. In [61] the authors implemented an intention recognition system using LMD. This information is then passed on to a robot which reacts to human intention. The authors in [62] developed a glove using Flex sensors and acquired the hand gesture data using camera. They also acquired the hand gesture data using LMD and performed comparison of the two sets of acquired data. They concluded that the LMD performance is better than that of the glove. An LMD and SVM based gesture detection system is proposed in [63]. They used an existing dataset which contains 1400 samples. In [64] the authors evaluated the performance of LMD. They collected data by giving the device to students who performed various tasks. The students provided a positive response for 79% tasks.

Table 2.2.1 is a summary of the undertaken review. The details include hardware, software, features, and limitations of the designed prototypes and products. The last column also includes information related to Sign Language Dataset (SLD).

Table 2.2.1: Summary of literature with focus on technology employed.

References	Hardware/Software	Summary
[48]	LMD, Android app	D-M to ND-M to D-M, Single SLD
[11]	PC, Harris algorithm	D-M to ND-M to D-M, Single SLD
[18]	Raspberry Pi	D-M to ND-M, Single SLD
[17]	LMD, LSTM Neural networks	D-M to ND-M, Single SLD
[16]	LMD, (GMM, HMM, DTW)	D-M to ND-M, Single SLD
[19]	PC, NN algorithms	D-M to ND-M, Single SLD
[5]	PC, Arduino, Gloves, Flex sensors	Calibration required for the custom-made product, D-M to ND-M, Single SLD
[20]	Arduino, Gloves, Sensors	Calibration required for the custom-made product, D-M to ND-M, Single SLD
[50]	PC, TensorFlow	D-M to ND-M, Single SLD
[51]	PC, Various ML algorithms	D-M to ND-M to D-M, Single SLD
[52]	PC, TensorFlow, CNN	D-M to ND-M, Single SLD
[36]	Android	D-M to ND-M, Multi SLD
[53]	Android	D-M to ND-M, Single SLD
[29]	PC, CNN	D-M to ND-M, Single SLD
[54]	Arduino, Gloves, Flex sensors	Calibration required for the custom-made product, D-M to ND-M, Single SLD
[32]	PC NN	D-M to ND-M, Single SLD
[55]	Gloves, Flex sensors, Voice recorder, Matlab	D-M to ND-M, Single SLD
[56]	PC, CNN	D-M to ND-M, Single SLD
[7]	CNN	D-M to ND-M to D-M, Single SLD
[31]	Gloves, Sensors, Arduino, KNN	D-M to ND-M, Single SLD
[14]	PC-based, PCA	D-M to ND-M, Single SLD
[8]	Arduino, Gloves, Sensors, KNN	Calibration required for the custom-made product, D-M to ND-M, Single SLD
[33]	PC-based, CNN	D-M to ND-M, Single SLD
[24]	Android application	D-M to ND-M, Multi SLD
[34]	PC, CNN	D-M to ND-M, Single SLD
[25]	Android	D-M to ND-M to D-M, Single SLD

2.3. Sign Language Dataset Review

In [52], the authors used the BSL dataset, containing more than 2500 images, to validate their prototype. The authors in [53] used a new method to detect an ArSL dataset. In [46], the authors used a word-level ASL dataset. Their selection of the dataset is based on the number of people who can use this sign language. Argentinian Sign Language (ArgSL) is processedin [41], where they used videos as input. The researchers used Turkish Sign Language (TSL) [45] and the CNN algorithm to process the dataset.

Chinese language speakers are the single largest group as compared to speakers of all other languages. The authors in [43] used Chinese Sign Language (CSL) with hundreds of categories within the dataset. Another word-level ASL dataset is used in [47]. In [42], a large-scale ASL dataset is presented and processed. Russian Sign Language (RSL) is used as input in [44]. In [22], the authors processed ArSL. A prototype was developed based on Arduino and a glove with sensors connected.

In [65] the authors developed an application to detect Portuguese Sign Language (PrSL). They created a game for school children to create awareness and highlight the importance of sing language. The authors used a COTS camera to capture sign language gestures. The authors created first dataset for Qatari Sign Language (QST) in [66]. This is an ongoing project where the size of the dataset is increasing. The authors created a virtual sign language interpreter and provided real-time conversion. Indonesian Sign Language (InSL) translation system is presented in [67]. The proposed system used COTS LMD, KNN and SVM. They reported accuracy of more than 93%.

In [68] the authors developed a system that detects Mayanmar Sign Language (MSL). The sign language is acquired using webcam. Korean Sign language (KSL) translation system is proposed in [69]. The authors used a dictionary to convert words into sign language. The Korean language dictionary is used an input which contains 12000 words. Chinese sentences are generated from Taiwanese Sign Language (TaSN) in [70]. The proposed system is used to detect sentences and converted into sign language.

In [71] the authors presented a system that detects Filipino Sign Language (FSL) using LSTM. They reported an accuracy of more than 95%. The authors proposed a system that detects Pashto Sign Language (PaSN) in [72]. They used CNN to process image data. The accuracy was 98%.

Table 2.3.1 is a summary of the different sign language datasets reviewed. This review is focused on the size of the datasets, how many people can use this dataset, and its complexity.

Table 2.3.1: Summary of sign language datasets reviewed.

	Sign Language Dataset	**Dataset size**
[52]	Bangladesh Sign Language	2660 images
[53]	Arabic Sign Language	50 signs
[46]	Word-level American Sign Language	More than 2000 words
[41]	Argentinian Sign Language	More than 3200 videos
[45]	Turkish Sign Language	226 signs
[43]	Chinese Sign Language	500 categories
[47]	Word-level American Sign Language	More than 10,000 videos
[42]	American Sign Language	10,000 signs and 25,000 videos
[44]	Russian Sign Language	164 lexical units
[22]	Arabic Sign Language	Not provided
s[65]	Portuguese Sign Language	Not provided
[66]	Qatari Sign Language	20,000 sentences
[67]	Indonesian Sign Language	1300 images
[68]	Mayanmar Sign Language	5000 images
[69]	Korean Sign language	12,000 images
[70]	Taiwanese Sign Language	1178 images
[71]	Filipino Sign Language	200 images
[72]	Pashto Sign Language	2500 images

2.4. Machine Learning Algorithm Review

In this section, a review of some machine learning algorithms is presented. A new dataset was created for supervised machine learning in [39, 40]. The dataset is used to suggest a solution to the faults found during electronic product manufacturing. The algorithms are implemented in LabVIEW [73]. The authors in [74] reviewed machine learning algorithms and used them to determine the performance of soldering stations.

In [75], the authors review machine learning algorithms. Their finding includes highlighting the limitations of some algorithms. The authors used Scikit, a Python [76] toolkit, to implement a machine-learning algorithm in [77].

2.5. Summary and Research Gap.

This section provides the conclusion of the literature review and highlights some limitations. In the research work reviewed in Section 2, the implementation is mostly meant for ND-M to understand sign language using either a PC-based or a standalone system. Some researchers created their devices using sensors and gloves and demonstrated their prototypes. It is important to understand that converting a prototype into a product is itself a huge task and requires a lot of work and research. These products will then have to go through a tedious validation and verification process. This means that it will take a lot of time for theseproducts to be ready for use.

The proposed system focuses on the limitations of existing systems in terms of providing an end-to-end system where both D-M and ND-M people can communicate in real time. The system uses a COTS device to acquire hand gesture signals, which means there will be no need to design and then validate custom-made hardware. This will allow the product to be available for use quickly. The other feature is that the system is not application dependent andcan be used in any environment where communication is required between D-M and ND-M people. The research is also focused on a low-cost system that is affordable and can be used with minimum training.
The majority of the work reviewed here is unidirectional, i.e., detection and conversion of hand gesture data into text and or voice. This is a limitation where the ND-M are unable to communicate with the D-M. The other limitation is the input sign language dataset, which in most cases, is one. This means that people who do not know that particular sign language are unable to benefit from the system. Some prototypes designed using gloves and sensorsrequire maintenance and calibration and are difficult to maintain consistency.

The system proposed in this research work provides a full-duplex system where both the D-M and the ND-M can communicate with each other. This system uses multiple sign language datasets, provides training, is easy to use, and has a low cost. Systems based on COTS devices like LMDs require considerably less effort in converting the prototype into a product. Affordability is also important, so low initial cost and no operational cost are essential featur

Chapter 3: Two-way Communication System based on Single Sign Language

3.1. Research Methodology

This section presents the research methodology of D-M-related research work. The research is carried out in two steps, which are shown in Figure 3.1.

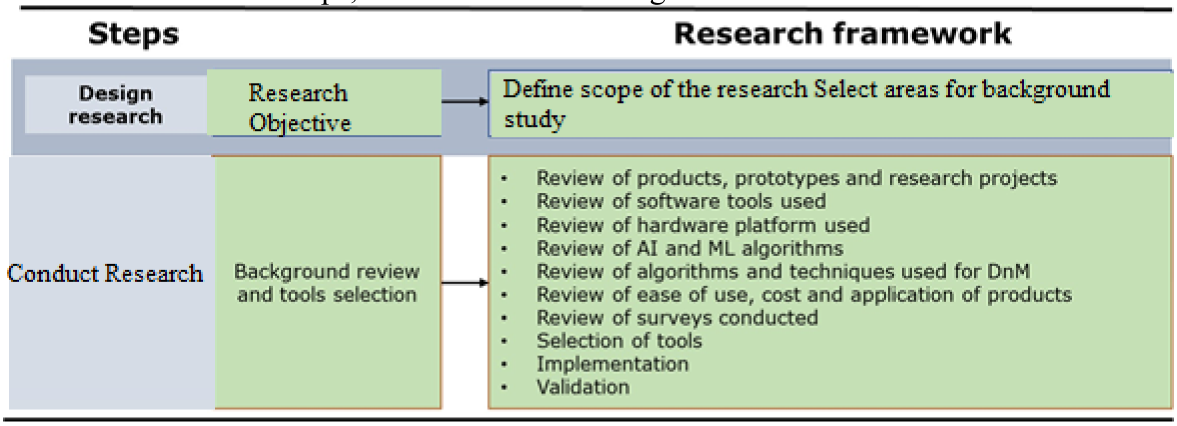


Figure 3.1. Research Methodology.

3.1.1. Design Research

The first task is to define the scope of this research and identify the activities to be carried out within the scope. The sequence of the activities is also defined. The first step is listing the areas where a review is required. Literature related to the area of this research is selected for review. The sub-category to review includes the software and hardware tools used, ease of use, how performance is evaluated, the algorithm, and techniques used.

3.1.2. Conduct Research

The second task is to conduct research and complete the tasks defined in the previous task. The first step is to collect existing research work within the scope of this research. The next step is to review some AI and ML algorithms and techniques. An extensive review is then carried out on the research work selected, which includes a review of the hardware and software platform, quality of the product or prototype, ease of use, and accuracy. After completing the review, tools are selected for progressing with this research. The last steps are implementation and validation.

3.2. Proposed System Block Diagram and Novel Features

In this chapter, the novel features of the proposed systems are presented, followed by a detailed description of the system.

3.2.1. Novel Features of the Proposed System

Figure 3.2 presents the six key features of this PC-based system. The proposed system is low cost and provides two-way communication between D-M and ND-M people. The system proposes a dataset for supervised machine learning to detect languages. The system uses a leap device t and a sound card for two-way communication. The proposed system provides a complete solution for communication between D-M and ND-M people. A unique feature of this system is its proposed dataset for detecting multiple languages. The PC software application is developed in LabVIEW.

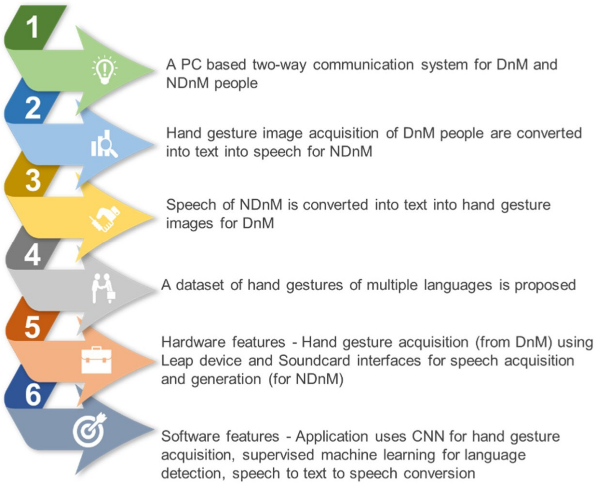


Figure 3.2. Novel features of the system.

3.2.2. Proposed System Block Diagram

The proposed system provides two-way communication between D-M and ND-M people. The hardware block diagram is shown in Figure 3.3. Hand gestures from D-M people are acquired using a Leap device, and the data are then sent to the PC application. The leap device captures the hand gesture data as an image, which is then processed to obtain the recognizable gesture. The data then goes through several activities implemented in the software and will be discussed later. The data are finally converted into speech, which is generated through sound card speaker output. ND-M people can listen to the speech and then respond through the speech, which is then acquired using PC microphone input. The acquiredaudio is then displayed as text on the screen for the D-M person to read.

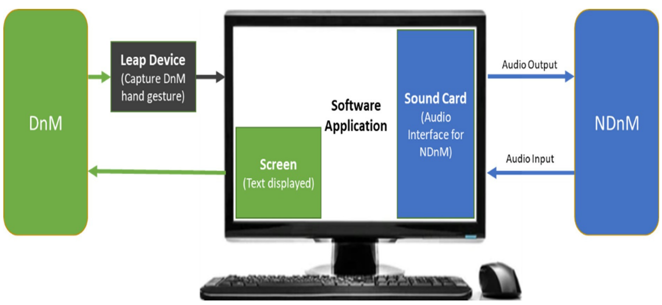


Figure 3.3. Hardware Interface Block Diagram—D-M/ND-M communication.

Figure 3.4 shows the various software building blocks. The gesture data are acquired and processed by the Leap device. The processing is based on MobileNet to detect gesture data. The data are then fed to the PC software application, which is developed in LabVIEW. The data are first processed using a supervised machine-learning algorithm that detects the language. This is done using a dataset created as part of this research work. The LabVIEW application then converts this into text that the ND-M person can read. The last step is to convert the text into speech, which is generated through a PC sound card. This completed the communication from D-M to ND-M. The ND-M person records the speech using a PC sound card microphone input. Speech is acquired, and then language is detected using supervised machine learning. The detected language is then converted into text, which is finally displayed for the D-M person to read.

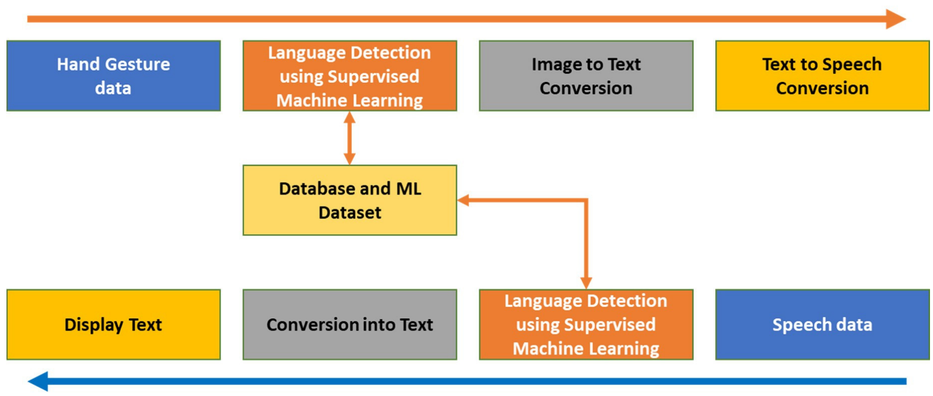


Figure 3.4. Software Block Diagram—D-M-ND-M communication.

Figure 3.5 shows the machine learning-based implementation through 8 steps. The system acquires hand gesture images using a Leap motion device, which is then processed, and the results are generated. The proposed system collects data that are used to improve system performance. The system also collects user information, which is very useful for improving the machine learning algorithm and implementation. Some parameters used for processing the image are listed in steps 2 and 3.

In figure 3.5, the process starts by providing relevant information to the machine learning algorithm through the input layer as shown by '1'. The processing is completed within the hidden layers and activities are marked by '2' till '7'. Step '2' is the initialization stage where the input image data goes through different tasks as listed. Step '3' is the processing stage where the image data is processed to detect certain features. Sign language dataset is detected in step '4'. The next step is to detect language. Currently there is only one language i.e., English as listed in step '5'. Data is stored in step '6' while the last step i.e., '7' in the hidden layer is for user data collection. The results are generated through the output layer as marked by step '8'.

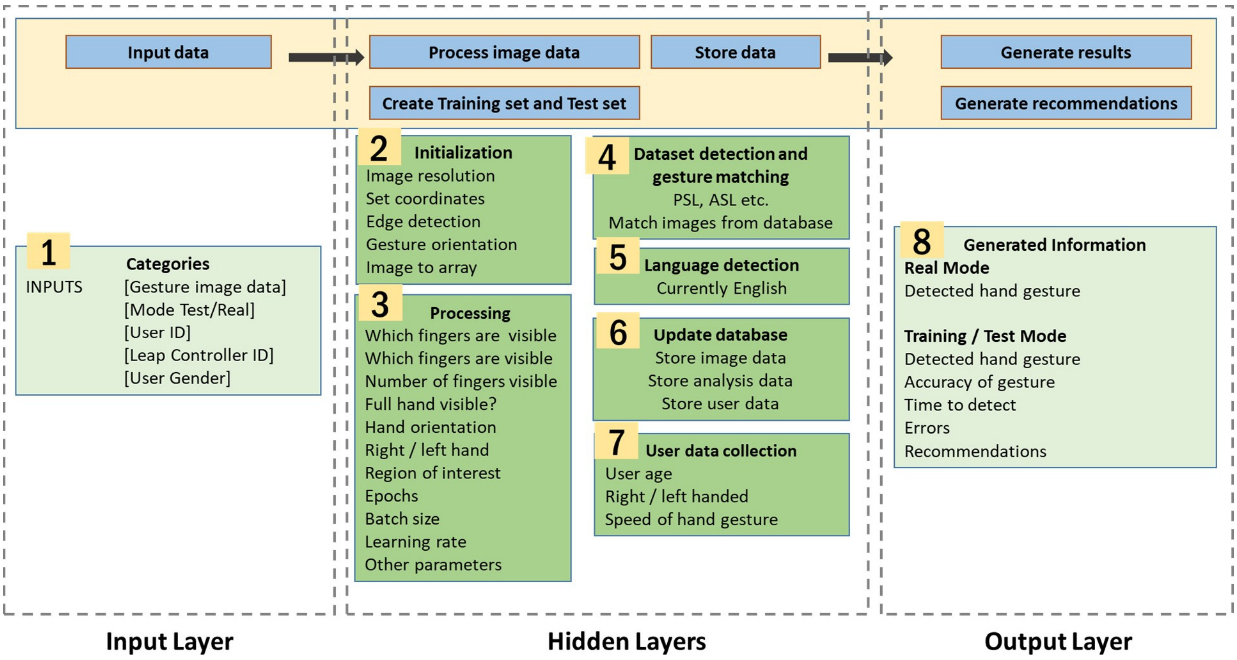


Figure 3.5. Machine Learning-based Actual and Training System Implementation.

Figure 3.6 shows a screenshot of the main software application. In this instance, the D-M person generated 5 gestures, i.e., 'H', 'E', 'L', 'L', and 'O', which are detected at a time. The software application created the word using these. Finally, the text is converted into audio and generated for ND-M people to listen to. In response, the ND-M person replied with the word 'GOOD', which is acquired through microphone input and converted from speech to text for the D-M person to read.

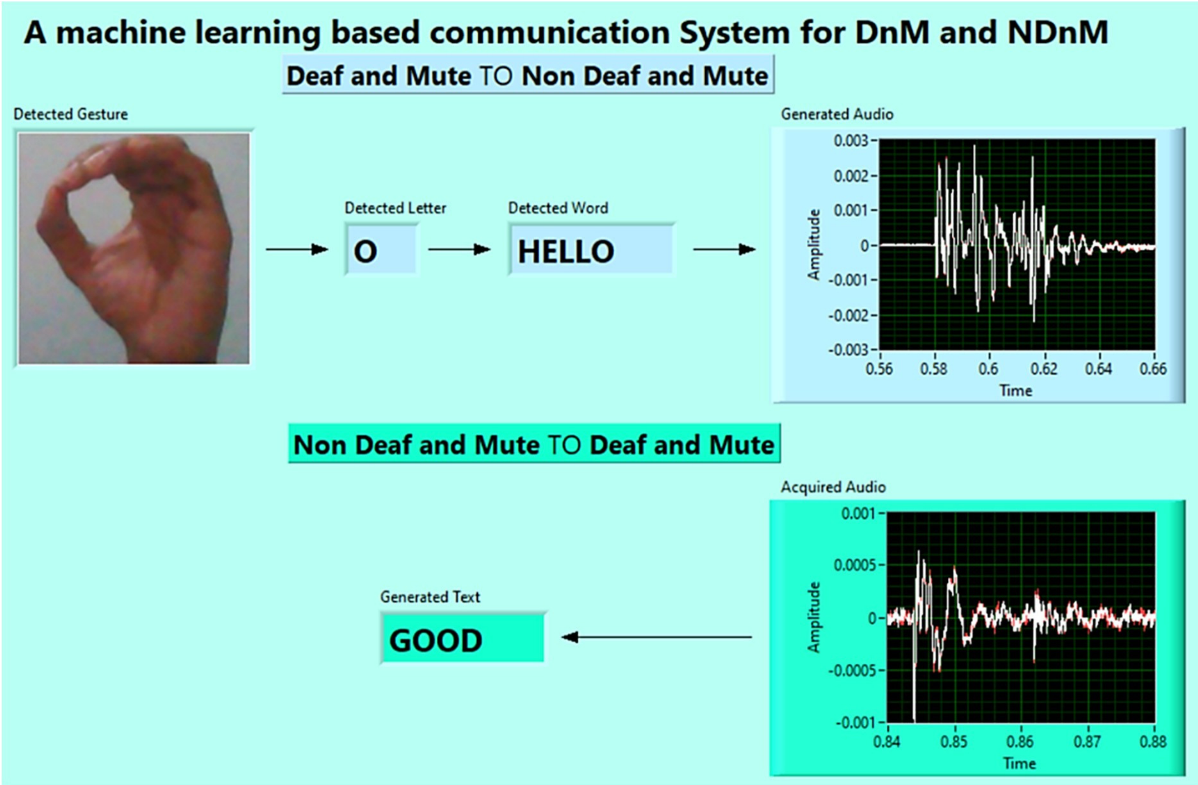


Figure 3.6. Software application screenshot.

3.3. Experimental Setup and Results

The validation of the proposed system and subsequent results are presented in this section. Section 3.3.1. lists the criteria and data used to validate the proposed system.

3.3.1. System Validation Dataset and Criteria

To validate the proposed system, ASL is selected mainly due to the availability of a full dataset, which is needed to create an effective training set. This dataset is used globally; thus, the number of end-users will increase. Other parameters considered are that this dataset canbe modified and redistributed, open source, use its own unique grammar, and how many people are using this dataset. An in-depth review of the sign language dataset with the aim of creating a new dataset is outside the scope of what is presented in this manuscript.

For any new proposed system, the first step is always to select known input test data and determine the expected output. Through this approach, the system can be validated, limitations can be understood, and accuracy can be defined. Once this is done, the secondstep is to expand the input test dataset and process it through the system. This process will help enhance the capabilities of this new system.

Using a known dataset is the key to validating the proposed system. A fully working and validated system will then open the door for further research by processing new datasets through the system.

Here, a validation strategy is defined in which some parameters are fixed to evaluate the proposed system. The first is to select a dataset for hand gestures. For this, the images are used from Kaggle [78]. The hand gesture images used for validation of the proposed system are of ASL. This process is used to review the proposed system for accuracy, speed, ease of communication, and effectiveness of the machine learning algorithm. The images used are in color and RGB form, although other forms can also be used.

Alphabets A to Z and numbers 0 to 9 are used for validation. Within the dataset, there are 1820 images of alphabets A to Z and 700 images of numbers 0 to 9. These images show different variants of these alphabets and numbers. The number of images is appropriate forthe evaluation of the algorithm. Images of some alphabets and numbers selected randomly are shown in Table 3.3.1.1. These images provide a variety of gestures that are detected and evaluated.

Table 3.3.1.1: A subset of alphabets and numbers from the Kaggle dataset

				
A	S	I	H	Z
				
0	1	2	5	8

The dataset is split into two groups. The first includes 70% of the data, which is used for training, while the remaining 30% is used for testing. This is a standard split for evaluating machine learning-based algorithms. In other combinations, researchers use 60% of the data for training, while 20% each is used for validation and testing, respectively.

For alphabet classification, 1274 images, i.e., 70% of 1820, are used for training, while the remaining 30%, i.e., 546 images, are used for testing. Similarly, for the evaluation of numbers, 70% of images, i.e., 490 out of 700, are allocated for training, while the remaining 210 images are used for testing.

3.3.2. Experiment 1 - Processing Quality of Acquired Gesture

In this experiment, different variants of one image of ‘A’ are selected, and hand gesture detection is processed. The gestures in all the images are identical, and only the quality is varied. Figure 3.7 shows the software screenshot showing the different qualities of the same image of gesture ‘A’. The image quality is changed to create five more images of gesture ‘A’, and then these images are processed using the image detection algorithm.

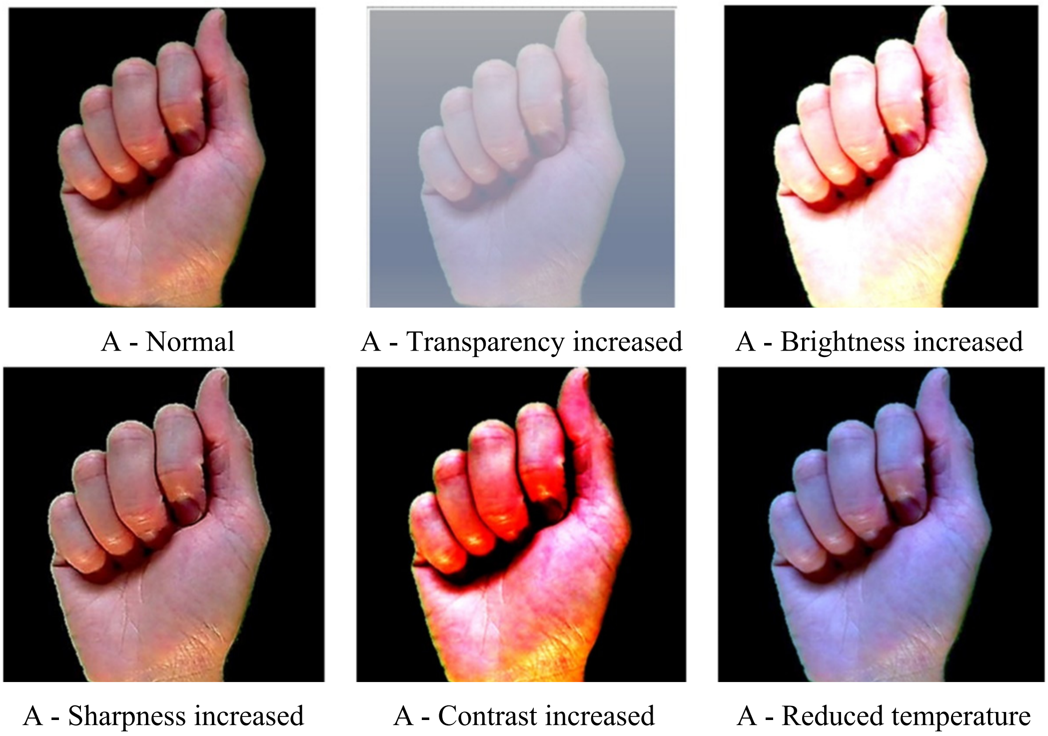


Figure 3.7. Hand gesture dataset for experiment 1.

Figure 3.8 provides a graphical representation of the findings. Each picture is run through the processing system a hundred times, and the accurate detection % is shown. The results of this experiment indicate that the accuracy of gesture recognition is higher than 93% across a variety of picture quality. We discover that the detection rate is an average of 96% when we look at all six photos combined. This is what we found out about it.

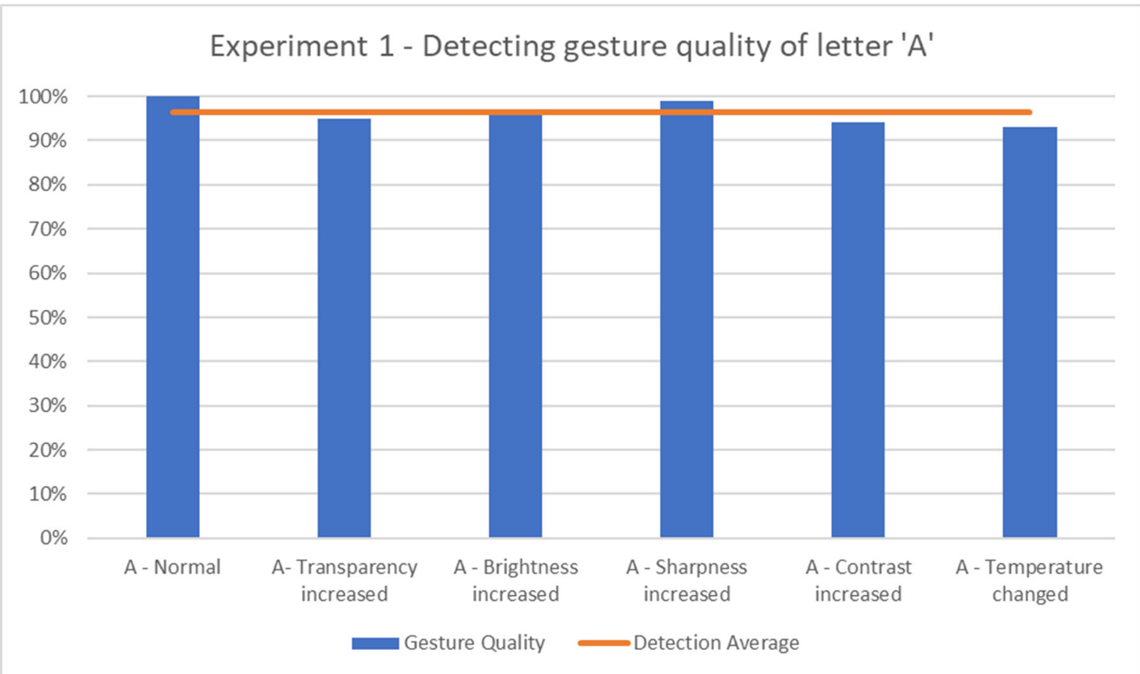


Figure 3.8. Quality detection accuracy of the algorithm.

3.3.3. Experiment 2 - Processing Variation in Gestures

In this experiment, two variants of three gestures, i.e., ‘H’, ‘I’, and ‘Y’, are used. The two gestures of each letter are different. Here, the algorithm is evaluated for detecting a gesture with a variation. Figure 3.9 shows images of the three gestures used for this experiment. The images in each row are of the same gesture but slightly different. The system performance is evaluated for this scenario by checking the accuracy of detection for these hand gestures.


H - Gesture 1 H - Gesture 2


I - Gesture 1 I - Gesture 2


Y - Gesture 1 Y - Gesture 2

Figure 3.9. Hand gesture dataset for experiment 2.

The results of the experiment are shown graphically in Figure 3.10. For this experiment, three different letters are selected with two slightly different gestures each. The detection accuracy is between 81% and 100%. As mentioned, in experiment 1, each image is processed 100 times. It is concluded from this experiment that the accuracy of some gesture variations isless than 90%. The overall average of correct detection in percentage is approximately 93%.

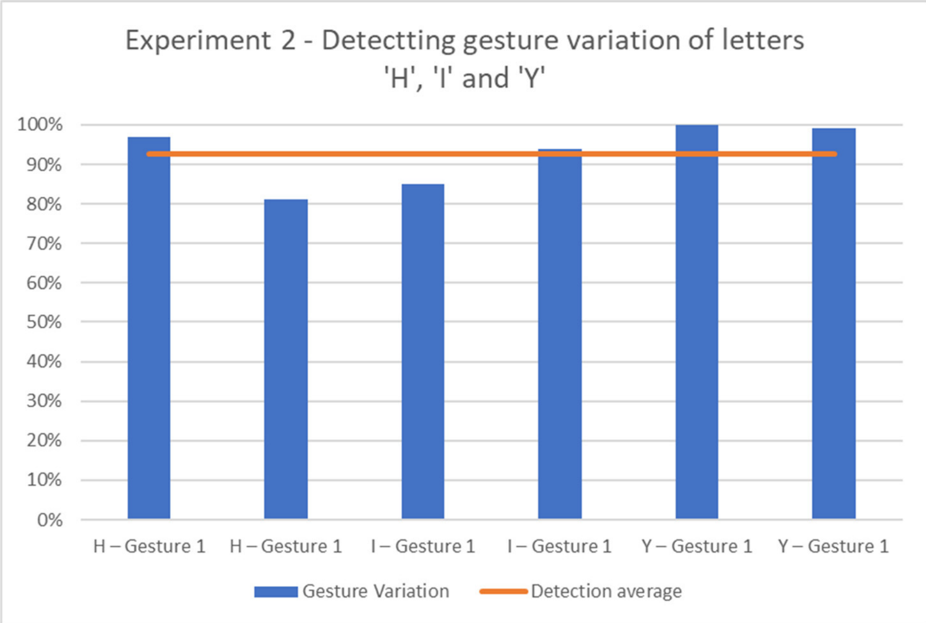


Figure 3.10. Variation detection accuracy of the algorithm.

3.3.4. Experiment 3 - Machine Learning Algorithm Repeatability

The accuracy of a model is a metric for determining which model is the best at recognizing correlations and patterns between variables in a dataset based on the input or training data. Accuracy is defined as the percentage of correct predictions, which is simply the ratio of correct guess or detection to the total number of predictions. This is presented in Equation(1).

$$\text{Accuracy} = \frac{\text{Correct Detection}}{\text{Total number of predictions}} \tag{1}$$

In this experiment, the repeatability of the algorithm is evaluated. The system is validated by processing the gestures of all letters and numbers. For validation, the most accurate gesture is selected and processed. The results are presented in Table 3.3.4.1.

Table 3.3.4.1. Accuracy using Mobile Net.

Letters and Number	Accuracy
A to Z	98.46
0 to 9	98.9

3.3.5. Experiment 4 - Algorithm Performance

The performance of the system is further evaluated using a confusion matrix. This is an N x N matrix, where ‘N’ denotes the number of target classes. The matrix compares the actual values and the values predicted by the machine learning algorithm. This provides a comprehensive picture of the performance of the machine learning algorithm. This also highlights errors in detecting gestures. The actual or expected values are shown in thecolumns, while the values predicted by the algorithm after detecting the gesture are listed in rows. The confusion matrix for letters 'A to Z' is shown in Figure 3.11, while the confusion matrix for numbers '0 to 9' is presented in Figure 3.12.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
J	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0.1	0	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	0.9	0.1	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
U	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0.9	0	0
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Z	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0.9

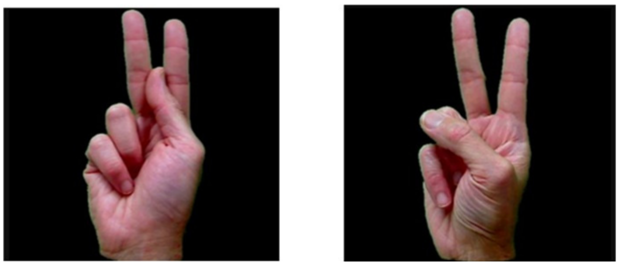
Figure 3.11. Confusion matrix of letters A to Z.

	0	1	2	3	4	5	6	7	8	9
0	1	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0
2	0	0	0.97	0.02	0	0	0.01	0	0	0
3	0	0	0	0.99	0	0	0	0	0.01	0
4	0	0	0	0	0.99	0	0	0	0	0.01
5	0	0	0	0	0	1	0	0	0	0
6	0	0	0	0	0	0	0.97	0	0.03	0
7	0	0	0	0	0	0	0	1	0	0
8	0	0	0	0	0	0	0	0	1	0
9	0	0	0	0	0.02	0	0	0	0	0.98

Figure 3.12. Confusion matrix of numbers 0 to 9.

3.3.6. Experiment 5 - Processing Similar Hand Gestures

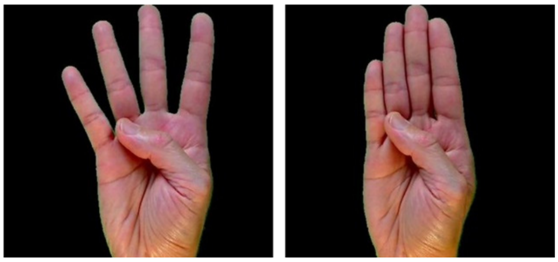
The reason for incorrect gesture recognition is due to some similar gestures. Similar gestures are shown in Figure 3.13. The accuracy of the detection of similar gestures is both dependent on the machine learning algorithm as well as on how correct the user's gesture is.



Gesture 'K' Gesture 'V'

Figure 3.13. Similar gestures Scenario 1.

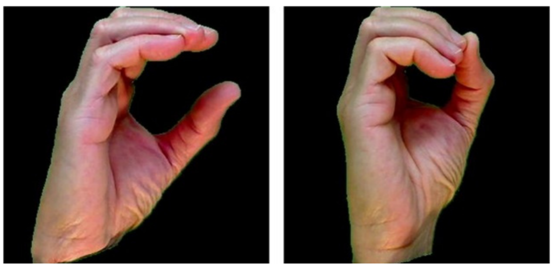
Figure 3.14 shows two different gestures with some similarities from ASL dataset. Both gestures have four fingers pointing upwards. There is a gap between fingers for gesture '4' while there is a no gap between fingers for gesture 'B'. System can detect wrong gesture in case of variation between gestures.



Gesture '4' Gesture 'B'

Figure 3.14. Similar gestures Scenario 2.

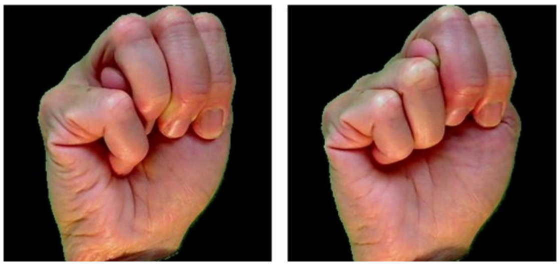
Figure 3.15 shows gestures 'C' and gesture 'Zero' from ASL dataset. The gestures have similarities and can be detected incorrectly if there is variation in the user gestures. If the thumb is close to fingers for gesture 'Zero' then this can be incorrectly detected as gesture'C'.



Gesture 'C' Gesture 'Zero'

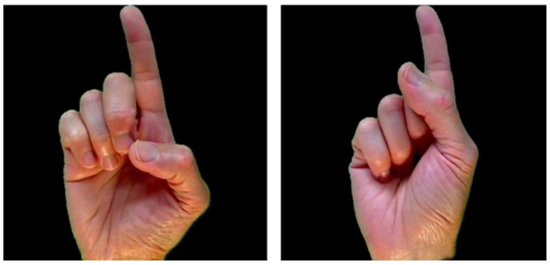
Figure 3.15. Similar gestures Scenario 3.

Figure 3.16 shows another scenario where two different gestures with similarities. As shown on the figure, the gestures 'M' and 'N' looks similar, and the difference is the position of the thumb.



Gesture ‘M’ Gesture ‘N’

Figure 3.16. Similar gestures Scenario 4.



Gesture ‘D’ Gesture ‘1’

Figure 3.17. Similar gestures Scenario

Figure 3.17 shows gestures ‘D’ and ‘1’ from ASL dataset. The main difference between the gestures is the position of thumb.The primary distinguishing factor among the gestures lies in the placement of the thumb

Chapter 4: Two-way Communication System for Deaf and Mute based on Multi Sign Language

4.1. Research Methodology

In this chapter, the research methodology is presented to explain the several steps which were carried out in this research. Figure 4.1 shows the block diagram with the steps and relevant details.

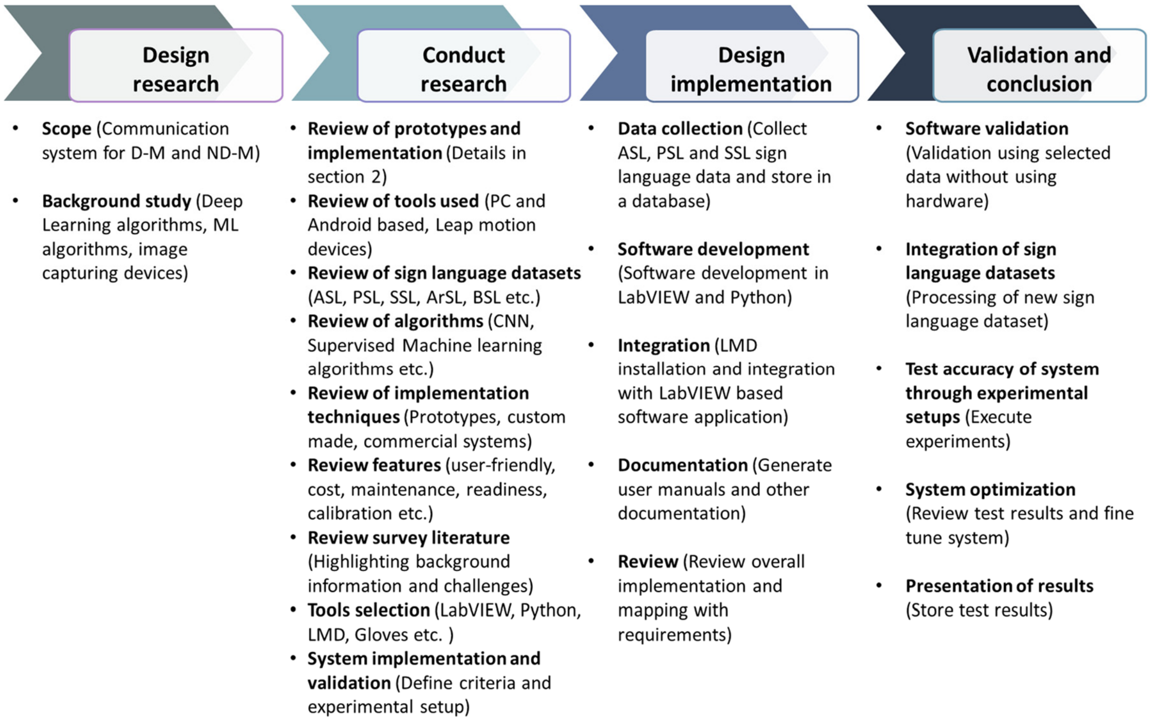


Figure 4.1. Research Methodology.

4.1.1. Design Research

In this step, two tasks are performed. The first task is to list the activities that are within the research scope. The scope of this work is to implement a two-way communication system where the D-M and ND-M can communicate with each other without the need to learn sign language through sign language interpretation and conversion to audio. The following task is to select certain areas to review the existing work, which in this case includes deep learning algorithms, ML algorithms, and image-capturing devices. These details are provided in chapter 2, literature review.

4.1.2. Conduct Research

In this second step, an in-depth review of the selected categories is carried out. The details are presented in figure 4.1.2. In this manuscript, some existing products and prototypes, which are developed to facilitate D-M communication, are reviewed. For this review, both the software and hardware aspects are considered, including the development platform, i.e., PC or Android, and hardware devices to capture hand gesture data. The literature review also covershow the existing systems are designed, whether based on COTS or custom-made. In the case of custom-made devices such as sensor-based gloves, the calibration and build process are also reviewed.

The next step is to select and review sign language datasets, machine learning algorithms, and deep learning algorithms. Within the review, the software applications are also considered, and LabVIEW is selected for software development. The final activity is to define experimental setups so that the proposed system can be validated. A detailed review is conducted in order to define how the different requirements for the proposed system can be mapped to the experiments and then validated. It is important to have an optimum number of experiments for validation as a means of avoiding repetition and duplication. The conclusion of the literature review is necessary to determine which parameters to use for implementing the proposed system.

4.1.3. Design Implementation

In this step, the tasks related to the implementation of the proposed system are considered. The features of the system are selected after conclusions drawn from the literature review.The design is implemented through five tasks, as listed in Figure 4.1. In the first step, the sign language data is collected and enhanced, i.e., some images are updated, and new images are added. The sign language datasets are selected, and information is stored in the database. The software, including the algorithm implementation, is done in LabVIEW. After completing the software development, hardware deployment, and integration tasks, the user manual and other documentation work are completed.

4.1.4. Validation and Conclusions

This is the final step, where the complete system validation is carried out. The validation activities include testing the updated language dataset, system response, processing times, and accuracy. The validated system is then tested for performance through different experimental setups, and the results are presented.

4.2. Proposed Communication System for D-M

In this section, the details of the proposed system are presented, including some key features.

4.2.1. Communication System Novel Features

Figure 4.2 shows the novel features of the proposed system, which provides full duplex communication between the D-M and the ND-M. This low-cost system is user-friendly and easy to install, with no operational cost to the end user. There is a training option available, which allows the features to be customized for individual users. The ML-based algorithm provides a continuous improvement option where new and higher quality data, i.e., hand gesture images, can replace an existing image. The proposed system supports American Sign Language, Pakistani Sign Language, and Spanish Sign Language. More languages can be added. Hand gestures data is acquired using Leap motion device (LMD), which is a Commerical-off-the-shelf (COTS) device and, unlike a glove, does not require maintenance or calibration. The image data is processed using the Convolutional Neural Network (CNN) algorithm.

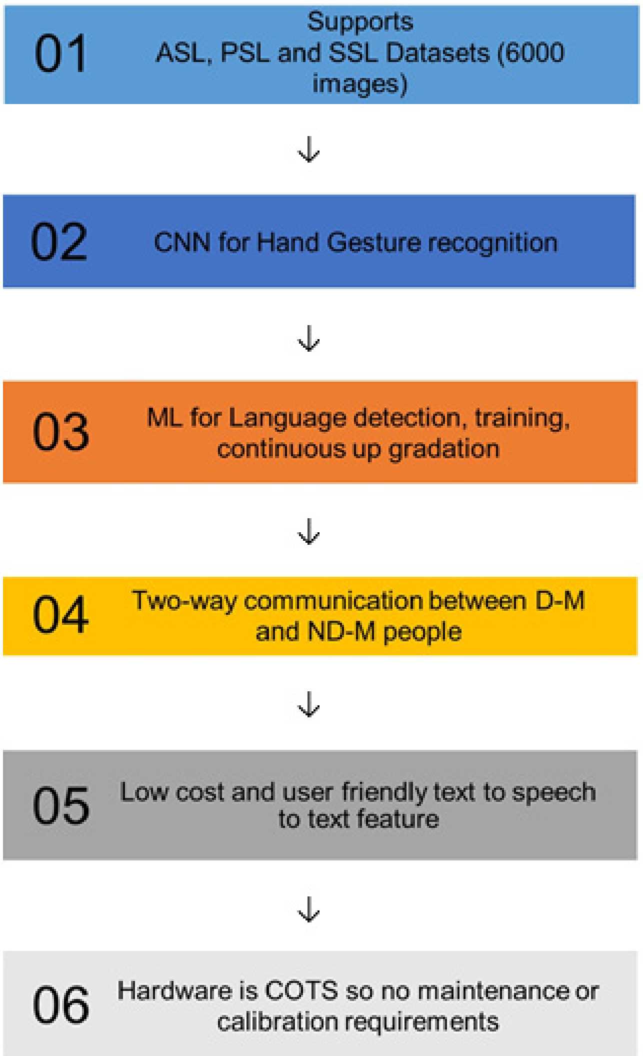


Figure 4.2. Novel features of the full duplex communication system.

4.2.2. Communication System Block Details

Figure 4.3 is a full duplex communication system between the D-M and the ND-M. The figure shows the implementation of the proposed system. The D-M person is provided with an interface where the hansd gestures are acquired using an LMD connected to the PC-based system. The LMD captures hand gestures as images and forwards them to the PC, where it is further processed. The details of the processing are discussed later. The system supports multiple sign language datasets, i.e., ASL, PSL, and SSL. The system is reconfigurable, and more datasets can be added in the future. The acquired image data is processed through multiple stages and is then converted into a voice message.

The ND-M person can listen to the voice message. The speech or voice data is generated through the PC sound card output. In response, the ND-M person records a voice message which is acquired by the PC using sound card input. The voice message is then processed by the software application and converted into text and hand gesture images. The D-M person is then able to read the text or see the images. Similarly, an ND-M person can also initiate a conversation.

The proposed system is low-cost, user-friendly, and requires less training time compared to other types of systems. These are the main features that make it easy and attractive for people to use this system. The sign language datasets selected for this system are based on the number of people using these sign language datasets. The selection of ASL, PSL, and SSL is based on the availability and size of datasets and the number of people using them. A small dataset means the size of the training dataset will also be small; hence the trained system will be less effective or accurate, while a large dataset results in higher accuracy but a slower response. Considering this, medium-sized datasets are selected.

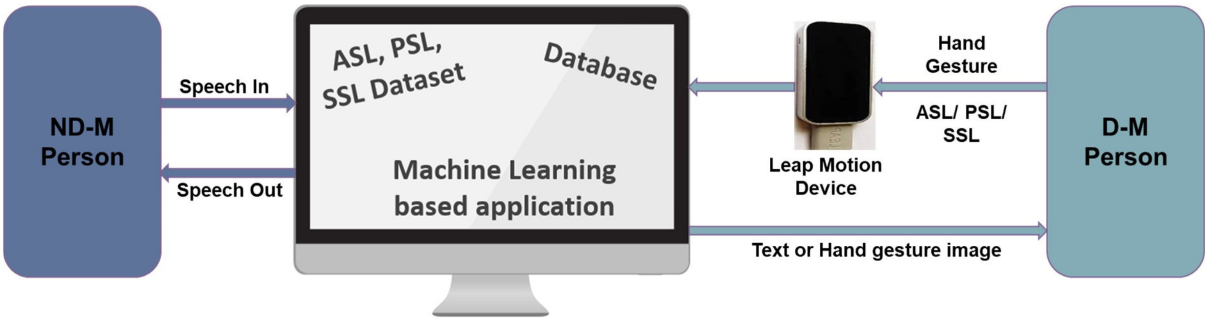


Figure 4.3. Block diagram of the communication system.

The system provides a training mode where the user can check the accuracy of the system while undergoing training. In this mode, the hand gestures of the D-M person are processed, and the detected results are displayed without going through the text to speech to textconversion. The D-M person can vary hand gestures to check for detection accuracy. In this mode, the system also updates the database by replacing existing images with higher-quality images. The ML-based process compares the newly acquired images with the stored images and then decides between replacing the existing image with the new one or storing the new image along with the existing image or not storing the new image at all. Having more image files means an improved dataset which will increase accuracy, but this also means the system will require more processing time. It is important to have a balance between these parameters.This decision is taken by the ML-based implementation.

The ML algorithm also reviews the user-stored data, which helps to increase processingspeed and accuracy. For example, if a user profile shows that the user only understands SSL, then the process will bypass the language detection step for this particular user. Similarly, the user can update the profile by adding more sign language datasets and other parameters. The system also maintains a performance record, both for the individual user and the overall system, using the stored user data and new data acquired through the training and normal modes.

Figure 4.4 shows the ML-based implementation. The block diagram shows the input, output, and hidden layers displaying various activities. The data is fed through the input layer, as shown by '1'. The input data includes both the hand gesture image data and the user data, which is stored as part of the user profile. In '2,' the user profile data is processed. The user can update the profile to reflect any changes. The next step, '3', is the training mode. This step is used when the training mode is selected. In normal mode, the ML algorithm can use some options from this step.

The language detection is done in step '4'. Currently, there are three datasets, but more can be added. The user-generated gestures are acquired by the proposed system, comparing them to the existing dataset, which is a combination of ASL, PSL and SSL. The system checks up to five gestures and then selects the language based on the results of the gestures matched to individual datasets of ASL, PSL or SSL. The acquired image goes through initialization in step '5.' In step '6,' the image goes through initial processing, where some features are detected and extracted. CNN algorithm is applied to further process the image data in '7.'The next two steps, i.e., '8' and '9,' are for data storage. The results are generated through theoutput layer as marked by step '10.'

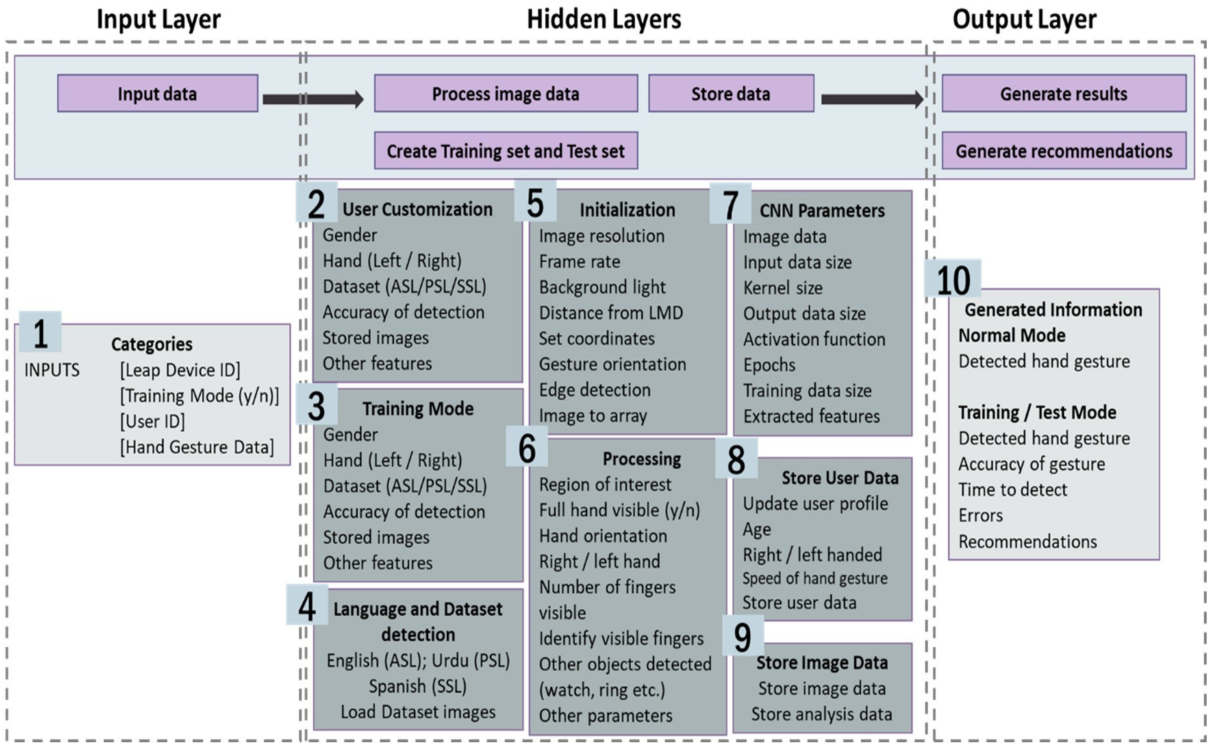
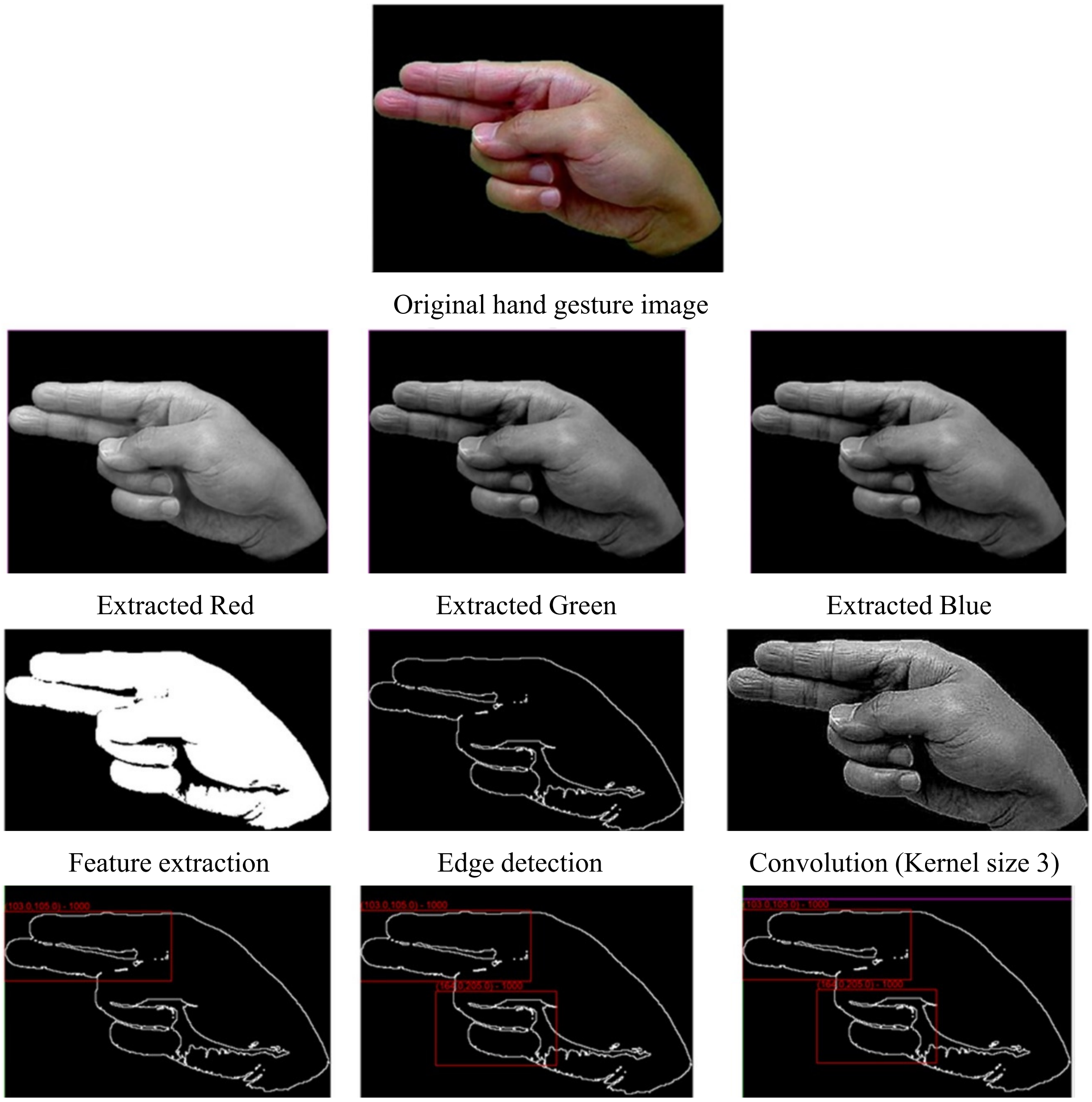


Figure 4.4. Machine Learning algorithm.

Figure 4.5 shows some graphical results of hand gesture image processing using the ML algorithm. The figure shows the original hand gesture and its extracted Red, Green and Blue (RGB) planes. The images in the third row show the feature and edge extraction of the original image for the particular gesture. The results of convolution using kernel size 3 are also presented here. The graphical results of detecting fingers using a region of interest and the pixel count of bright objects, i.e., the borders, are also shown.



Original hand gesture image

Extracted Red | Extracted Green | Extracted Blue

Feature extraction | Edge detection | Convolution (Kernel size 3)

Detect fingers (Region of interest) | Pixel count (3.1% of bright object)

Figure 4.5. Examples of processed images using the ML algorithm.

4.3. Experimental Setup and Results

In this section, the proposed system is validated using eight experiments that are specifically defined for this research. It is important to validate the key features and understand the limitations of the system. This is achieved using the experiments presented in this section.

4.3.1. System Validation Criteria

The communication system is validated through two stages. The first stage is to select the sign language dataset, which in this case are ASL, PSL, and SSL datasets. The second stage is to select a subset of these datasets. It is important to use a known input dataset and a predicted output result. This will help in understanding the system accuracy and limitations ofthe proposed system. Once the system is validated using these criteria, the full dataset is applied to further determine the performance of the system.

It is important to use a known set of images for initial validation. Outside the scope of this work is a study of how sign language datasets are created and validated.

To validate the proposed communication system, the ASL dataset images are used [78]. These are color images of hand gestures. The SSL dataset used for validation is available online [79]. The PSL dataset is created by recording images of letters and numbers. A reference to PSL is also available from [80].

Letters and numbers from the three sign language datasets are used for validating the communication system. For this research work, 70% of the hand gesture images from the datasets are used for training, while the remaining 30% are used for testing. The split of 70% and 30% is a good ratio for training and testing. However, a different ratio can also be used. As an example, approximately 1800 images of letters and 700 images of numbers from ASL are used, which are then split into the ratio mentioned before. Figures 4.6 – 4.8 show subsets of ASL, PSL, and SSL dataset images.

Figure 4.6 shows five numbers and letters from ASL, while in Figure 4.7, eight letters from PSL are shown.

.

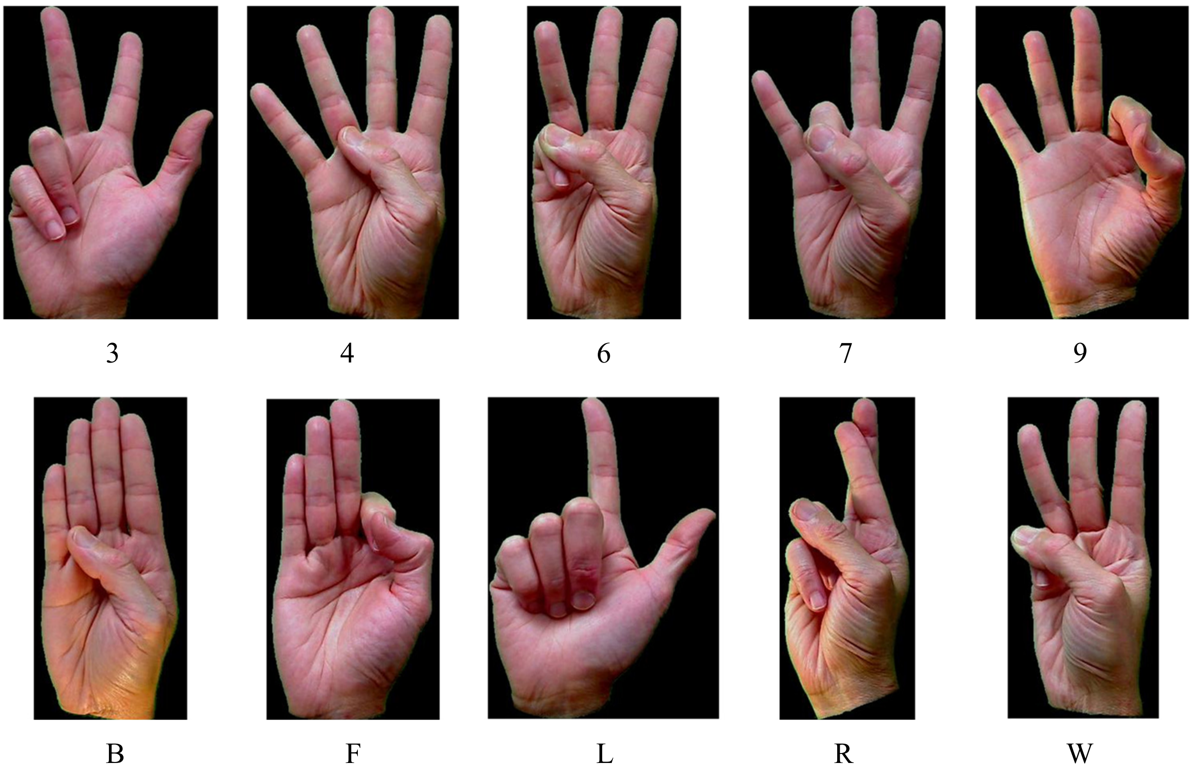


Figure 4.6. A subset of the American Sign Language dataset.

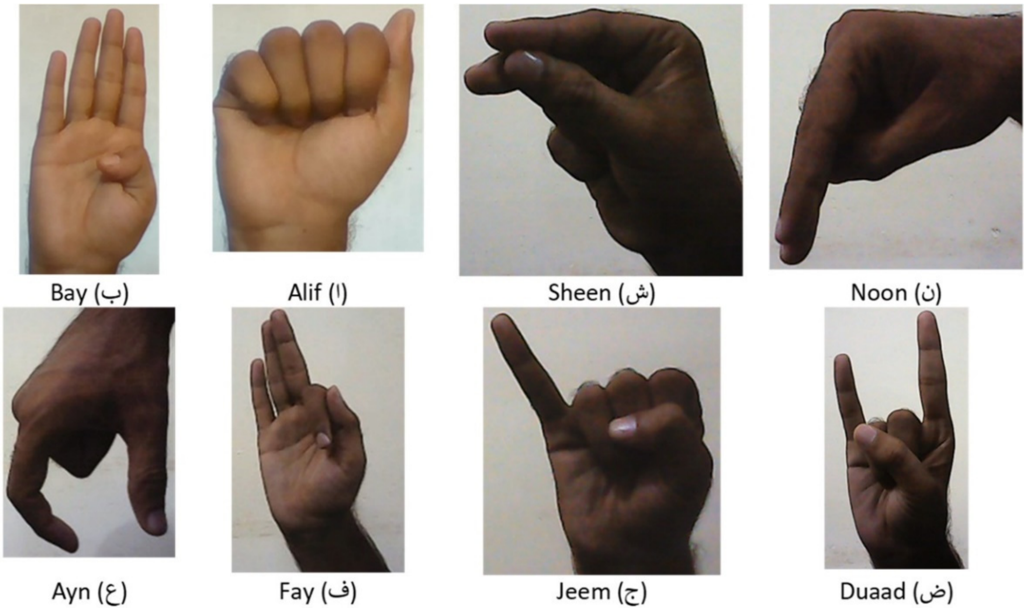


Figure 4.7. A subset of the Pakistani sign language dataset

In Figure 4.8, six letters from SSL are shown. The dataset used by the proposed system includes approximately 6000 hand gesture images from the three individual datasets.

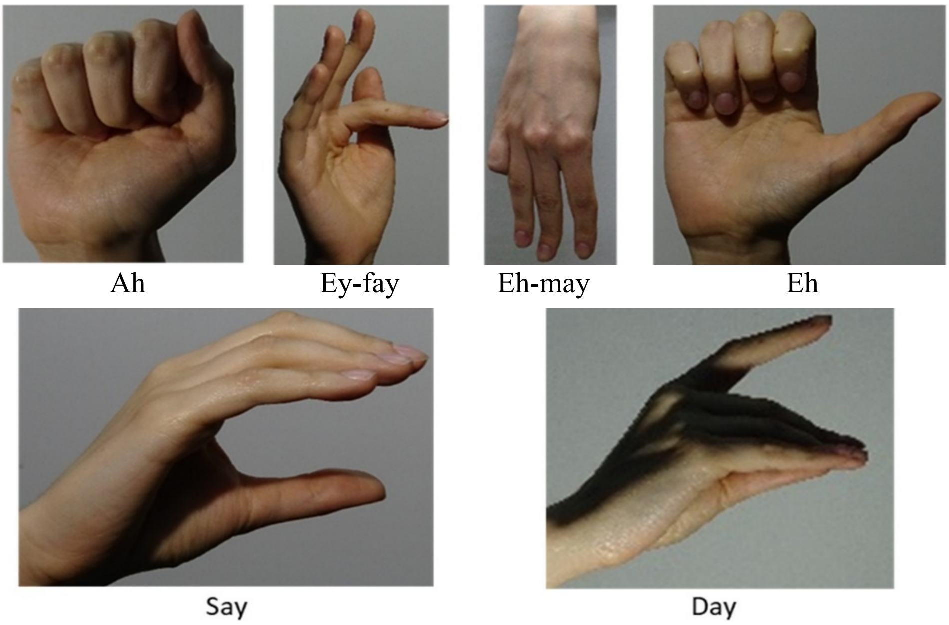


Figure 4.8. A subset of the Spanish sign language dataset.

4.3.2. Experiment 1 - Proposed Communication System Accuracy

In this experiment, the accuracy of the detection of hand gestures is presented. The formula for accuracy is presented in Equation (2). The accuracy is presented in percentages in Table 4.3.2.1. The accuracy values are also used to determine the performance of the algorithm.

$$\text{Accuracy} = \frac{\text{Correct Detection}}{\text{Total number of tries}} \tag{2}$$

In this experiment, the repeatability of the algorithm is evaluated. The system is validated by processing the gestures of all letters and numbers. For validation, the most accurate gesture detected by the system is selected and processed. The results are presented in Table 4.3.2.1

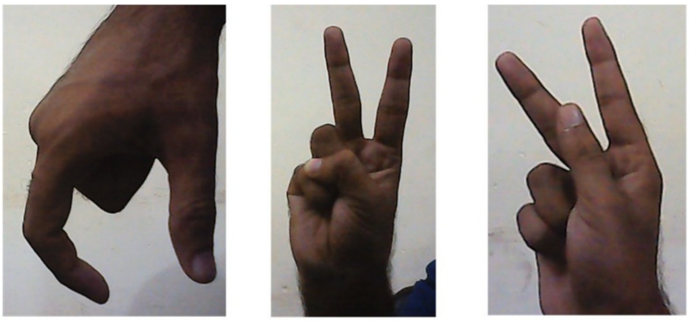
the accuracy is listed for letters and numbers of the three individual sign language datasets. The table also lists the size of subsets used for the training and testing of the system. It is also observed that the accuracy of the proposed system depends on the accuracy of the hand gestures, i.e., how accurately the gesture is made.

Table 4.3.2.1: Accuracy of detection.

Dataset	**% Accuracy (Letters)**	**% Accuracy (Numbers)**	**Train Dataset (70%)**	**Test Dataset (30%)**
ASL	98.5	98.9	1750	750
PSL	93.1	92.7	1015	435
SSL	94.3	93.8	1330	570

4.3.3. Experiment 2 - Processing Individual Hand Gestures

In this section, hand gesture detection results are presented. In Figure 4.9, three letters from the PSL dataset are randomly selected for this experiment and processed through the proposed system. The hand gesture detection algorithm acquires the hand gesture and compares it with the images from the dataset stored in the database. The results show that the detection can exceed 96%. Depending on the input hand gesture, more than one image from the database can be matched, and the one with the highest percentage is selected.



Letter 'Ayn (ع)' 96% Letter 'Rray (ڑ)' 99% Letter 'Kaaf (ک)' 96%

Detection accuracy of letters from PSL Dataset

Figure 4.9. Processing hand gestures (PSL dataset).

Figure 4.10 shows the accuracy and loss graphs for the PSL dataset. The graphs include both actual and test results.

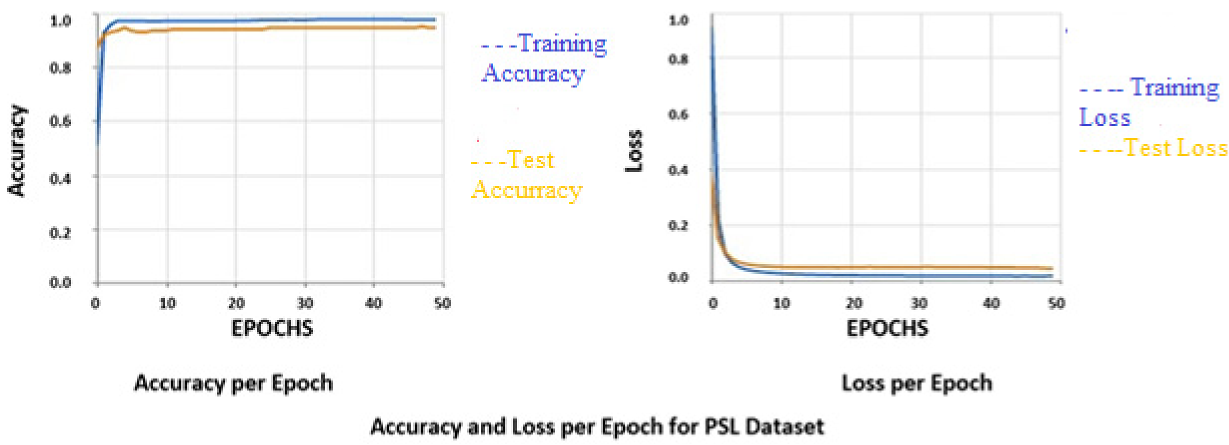


Figure 4.10. Hand gesture detection results for PSL dataset.

Figure 4.11 shows the hand gesture detection results for the SSL dataset. The results are encouraging and show high accuracy. In this figure, three random gestures are selected from SSL, and the detection accuracy is shown. It is observed that the accuracy of the captured hand gesture is high when it is unique among the stored hand gesture images. As an example, if two different hand gesture images look similar, then the detection accuracy of these will be slightly lower.

Letter ‘Ah’ 93% | Letter ‘Bay’ 95% | Letter ‘Say’ 96%

Detection accuracy of letters from SSL Dataset

Figure 4.11. Processing hand gestures (SSL dataset).

4.3.4. Experiment 3 - Processing Hand Gestures Image Quality

In this section, the experiment is carried out by changing the quality of the hand gesture image and then processed through the proposed system. Figure 4.12 shows the four variants of a letter from the PSL dataset. All the images are correctly detected, thus confirming that the proposed system is able to detect the correct gesture for varying image quality. In this experiment, the image quality of the acquired hand gestures varies, and the results show that the system is able to detect the letters correctly. This is an important result since the acquired image quality can vary due to background color, sharpness, background texture, etc.

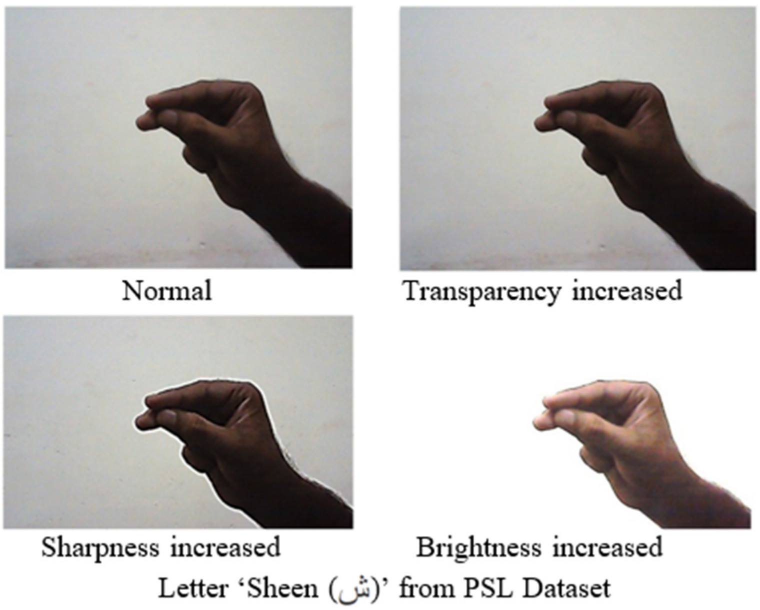


Figure 4.12. Processing images of different quality for the same hand gesture.

4.3.5. Experiment 4 - Processing Variations in Same Hand Gestures

In this section, the variation in hand gestures is processed through the proposed communication system. In Figures 4.13 – 4.15, two variants of letters from the PSL dataset are processed, and the detection accuracy is calculated. In this experiment, three letters from PSL are selected, and the hand gestures of these letters are varied slightly and then processed through the system. The results show that the accuracy is between 80% and 100% for the firstletter in Figure 4.13. This variation is due to the distance between the hand and the LMD. A part of the index finger of the image on the left is outside the camera vision.

Detection accuracy: 89% Detection accuracy: 100%

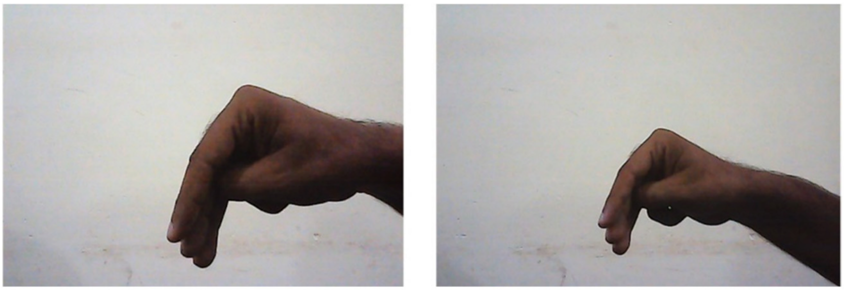
Letter 'Zoyay (ظ)' from PDL Dataset

Figure 4.13. Processing variation gesture scenario 1.

A similar experiment is repeated in Figure 4.14, where the accuracy is 100% due to minimal variation between the images. The results shown in Figure 4.15 are similar to that in Figure 4.13. Again, the low accuracy for the image on the right side in Figure 4.15 is due to the distance of the hand from the LMD as well as the variation in the gesture.

Detection accuracy: 100% Detection accuracy: 100%Letter 'Laam (ل)' from PDL Dataset

Figure 4.14. Processing variation in same hand gesture scenario 2.

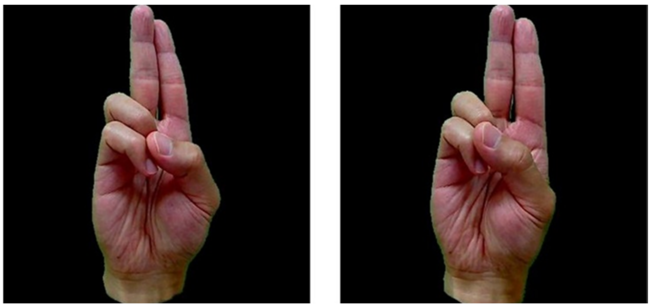


Detection accuracy: 100% Detection accuracy: 83%

Letter ‘Meem (م)’ from PDL Dataset

Figure 4.15. Processing variation in same hand gesture scenario 3.

In Figure 4.16, a letter from the ASL dataset is shown with a detection accuracy exceeding 90%. The accuracy for the two images is high for the letter selected from the ASL dataset, mainly because the distance between the hand and camera for both images is similar, and the variation between the two gestures is minimal.

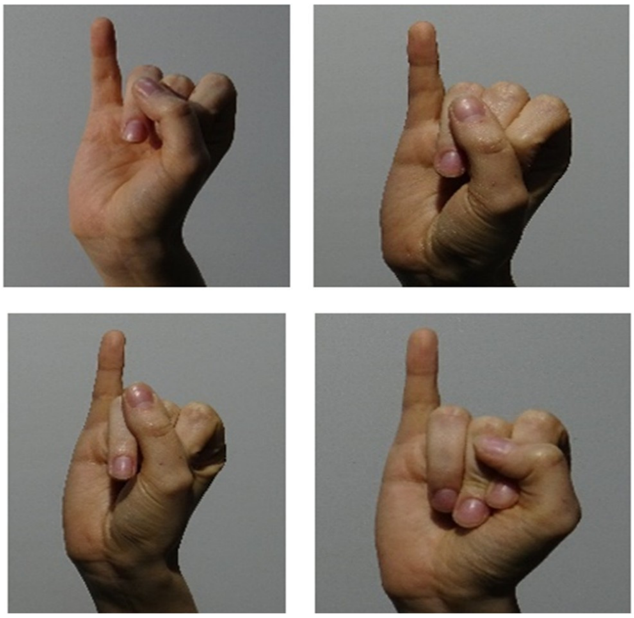


Detection accuracy: 94% Detection accuracy: 98%

Letter ‘U’ from ASL Dataset

Figure 4.16. Processing variation in same hand gesture scenario 4.

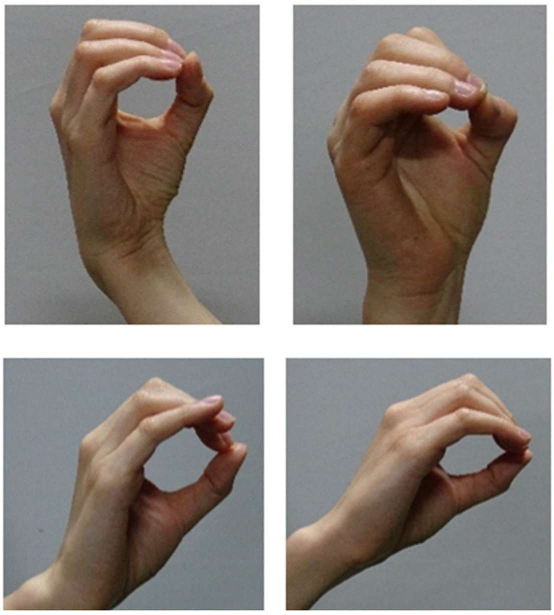
In Figure 4.17, letter ‘I’ is selected from SSL dataset. There are four variations of this letter shown in this figure. The average detection accuracy of all the gestures is more than 95%. The variation between the four gestures is minimal and the detection is correct.



Detection accuracy: 95%
Letter ‘I’ from SSL

Figure 4.17. Processing variation in same hand gesture scenario 5.

In Figure 4.18, letter ‘O’ is selected from SSL dataset. The four variations of this letter include change in orientation as shown in this figure. The average detection accuracy of all the gestures is more than 93%. The variation between the four gestures is minimal and the detection is correct.

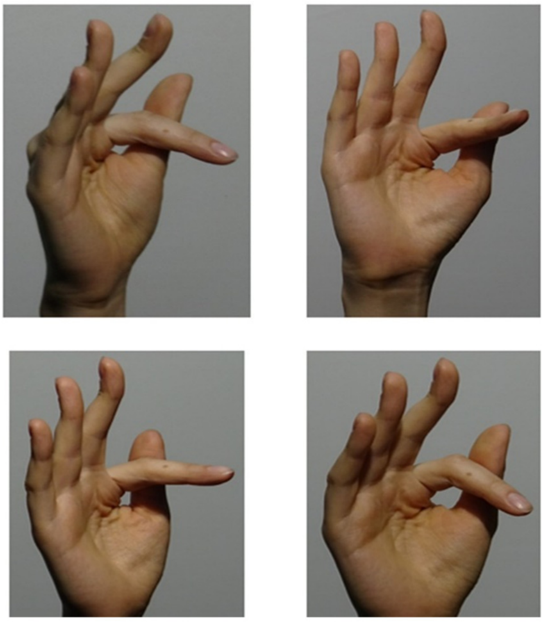


Detection accuracy: 93%
Letter ‘O’ from SSL

Figure 4.18. Processing variation in same hand gesture scenario 6.

In Figure 4.19, letter ‘F’ is selected from SSL dataset. The gesture is shown through four slightly different variations. The average detection accuracy of all the gestures is more than

92%. The main detection area for this gesture is the crossing of one finger with the thumb while three other fingers pointing upwards.



Detection accuracy: 92%

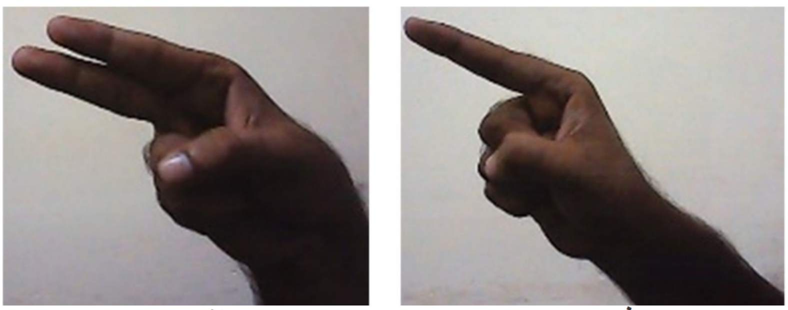
Letter ‘F’ from SSL

Figure 4.19. Processing variation in same hand gesture scenario 7.

4.3.6. Experiment 5 - Processing Similar Hand Gestures

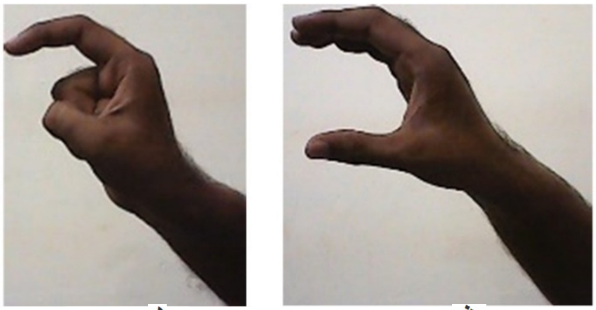
Some hand gestures are similar, although not identical, and consequently, there is a possibility of wrong detection. In this section, results of similar hand gesture detection are presented. Figures 4.20 and 4.21 show the detection results of two sets of letters from thePSL dataset.

In Figure 4.20, two hand gestures from the PSL dataset are compared. There is a significant difference between the two gestures, as there is only one finger used for the gesture on the right but two fingers for the gesture on the left. The system, in this instance, is able to correctly detect the two letters. The experiment is repeated for two more letters from PSL in Figure 4.21, which are different, and the system is able to detect these letters correctly.



Hey (ہ)   Zay (ز)

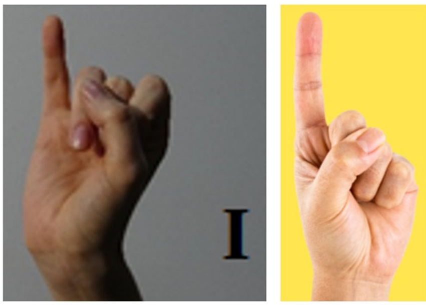
Figure 4.20. Processing similar gestures (PSL dataset) scenario 1.



Daal (د)　　Say (ث)

Figure 4.21. Processing similar gestures (PSL dataset) scenario 2.

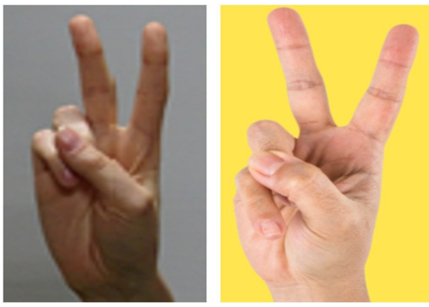
In Figure 4.22, two hand gestures from the SSL dataset are compared. These are letter ‘I’ and number ‘1’. It should be noted that different fingers are used for these gestures but still there is an element of uncertainty when detecting the gestures. This uncertainty means the accuracy is reduced slightly, but the proposed system is able to detect the gestures correctly.



Letter ‘I’　　Number ‘1’

Figure 4.22. Processing similar gestures (SSL dataset) scenario 3.

The same behavior occurs with letter ‘U’ and number ‘2’ in SSL. Although the accuracy is reduced slightly, the systems is able to detect the gestures again correctly.



Letter ‘U’　　Number ‘2’

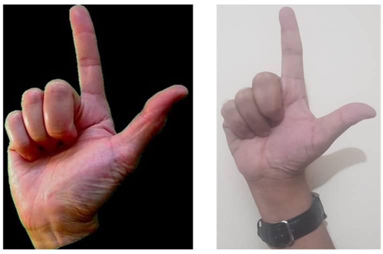
Figure 4.23. Processing similar gestures (SSL dataset) scenario 4.

4.3.7. Experiment 6 - Detecting Other Objects

This experiment focuses on other objects that can be visible while detecting hand gestures. The proposed system has a unique feature where these objects are also detected, and their information is ignored in some cases so that the correct hand gesture can be recognized. Figures 4.24 – 4.26 show images with and without other objects. A unique feature of this system is validated using this experiment. Here the user is wearing a watch in Figures 4.24 and 4.25 and a ring in Figure 4.26. It is observed that the system detection accuracy can vary if these objects are also captured with a hand gesture. Therefore, the images of objects like watches, rings, etc., are also stored in the database, enabling the proposed system to detect these objects and ignore them.

ASL Letter 'A'

Figure 4.24. Detecting other objects—wristwatch scenario 1.



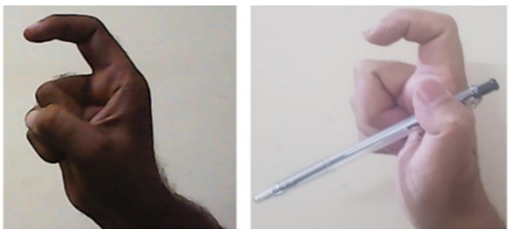
ASL Letter 'L'

Figure 4.25. Detecting other objects—wristwatch scenario 2.

PSL Letter 'Noon ( )' ن

Figure 4.26. Detecting other objects—ring.

Figure 4.27 shows another example of an object captured with a hand gesture. The proposed system is unable to correctly detect the hand gesture here. In this experiment, the object, which is a pen, is detected by the system, but the system is unable to detect the hand gesture correctly. This is due to the shape of the object and also how the user is holding it.



PSL Letter 'Daal (د)'

Figure 4.27. Detecting other objects—pen.

4.3.8. Experiment 7 - Algorithm Performance

In this section, the performance of the PSL dataset is evaluated and presented using aconfusion matrix. This matrix shows the results for each letter compared with all other letters.The matrix is used in evaluating the performance of the algorithm and highlighting anyerrors. The confusion matrix for letters within the PSL dataset is shown in Figure 4.28. Thisis an important metric for validating this type of system.

For clarity, the PSL set is :

ا	ب	پ	ت	ٹ	ث	ج	چ	ح	خ	د	ڈ	ذ	ر	ڑ	ز	ژ	س	ش	ص	ض	ط	ظ	ع	غ	ف	ق	ک	گ
ل	م	ن	و	ہ	ی	ے	ء																					

Add Figure no and add alphbets for ASL and SPL

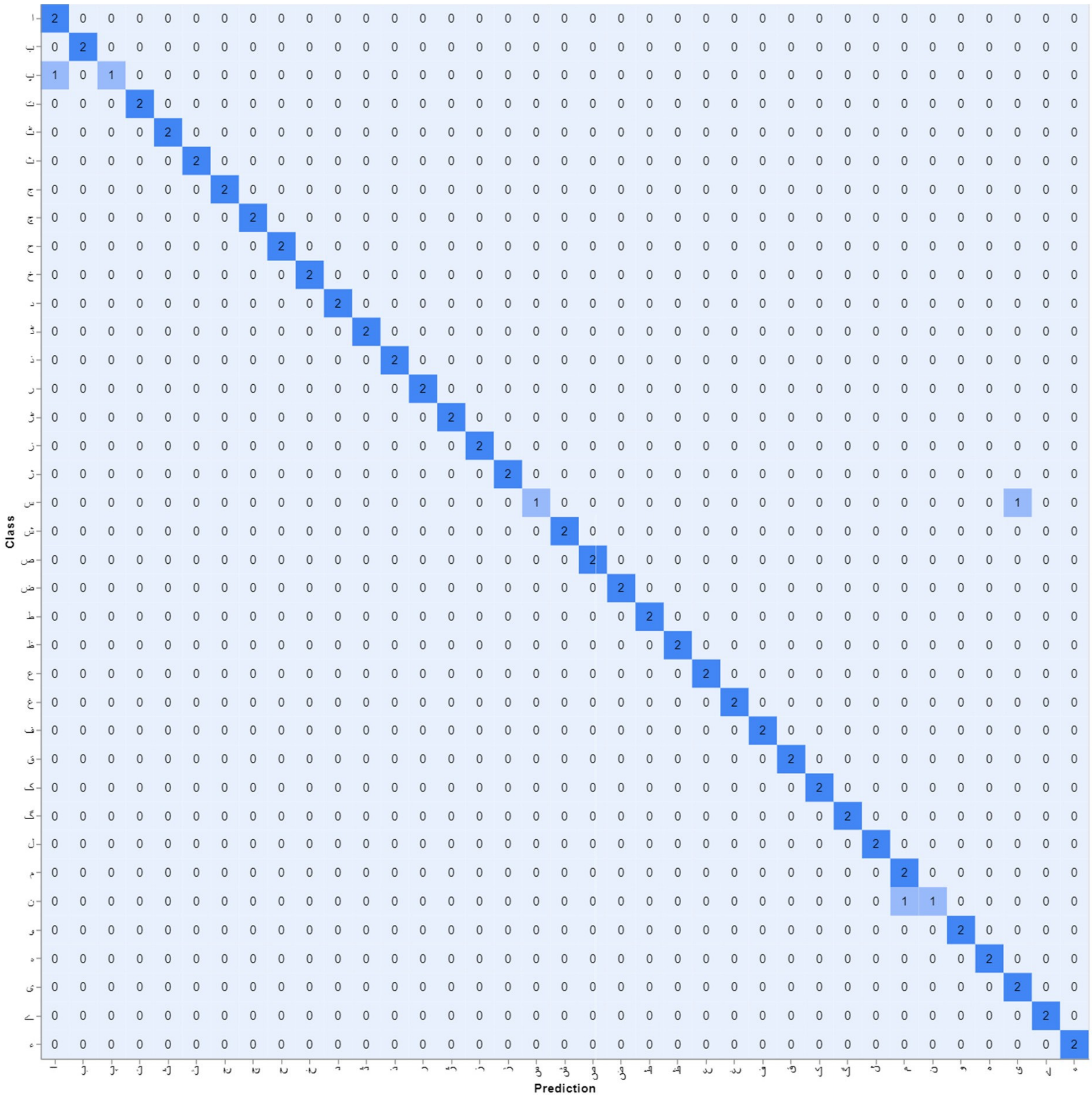


Figure 4.28. Confusion matrix of letters within PSL dataset.

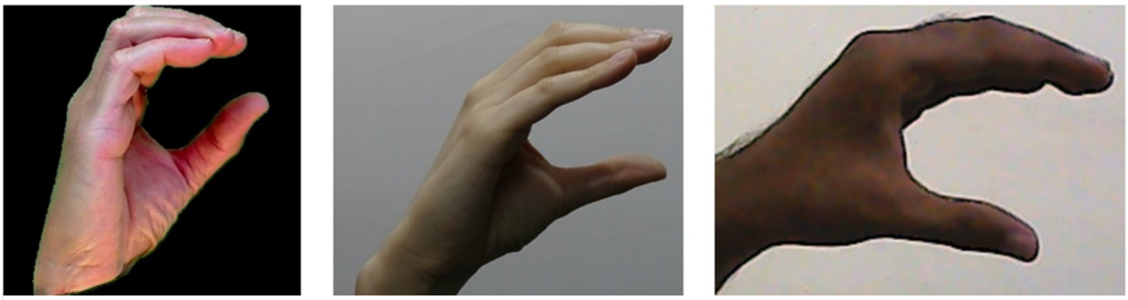
4.3.9. Experiment 8 - Creating a New Dataset

In this experiment, the similarities between the three sign language datasets are presented. In Figure 4.29, three letters, one each from ASL, SSL, and PSL, are shown. These letters have similar hand gestures. This approach can be used to create a new sign language dataset which can be a combination of multiple sign languages. This figure provides insight into the process of creating a new dataset. In the normal validation process, a confusion matrix is created, in which each letter within the dataset is compared with the rest of the letters, and accuracy is calculated. The figure shows a scenario where the gestures taken from ASL, PSL, and SSL are identical. In this case, the system is unable to detect the correct input sign language based on just one hand gesture.

ASL Letter 'A' SSL Letter 'Ah' PSL Letter 'Alif (ا)'

Figure 4.29. Similarities between different sign languages scenario 1.

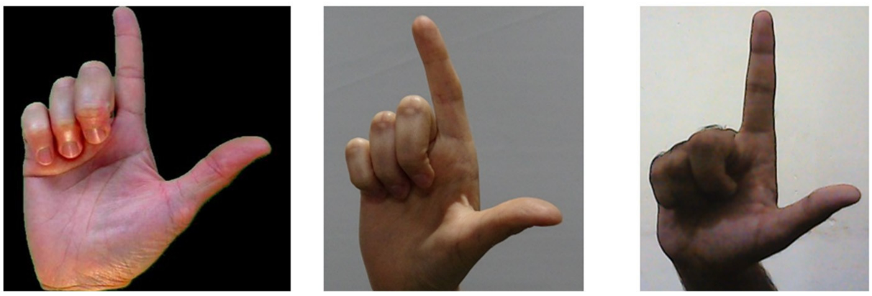
Figure 4.30 shows three letters, one each from ASL, SSL, and PSL. As shown in the figure that the three signs look similar. Comparing similar letters is important for accurately detecting correct sign language.



ASL Letter 'C' SSL Letter 'Say' PSL Letter 'Se (ث)'

Figure 4.30. Similarities between different sign languages scenario 2.

Figure 4.31 shows another example of similar gestures from the three sign languages. As part of auto detection of sign languages, the similar gestures are compared and one language is selected based on the best result.



ASL Letter 'L' SSL Letter 'ele' PSL Letter 'Lam (ل)'

Figure 4.31. Similarities between different sign languages scenario 3.

4.1.1. Experiment 9 - Creating Words

Figure 4.32 shows how the algorithm process individual gestures to create words. In this example a word 'in theWENT' is acquired by detecting ASL dataset.

Figure 4.33 shows a word is created from gestures detected from PSL dataset. The word is pronounced as 'Salaam' in Urdu. This word is used to greet someone.

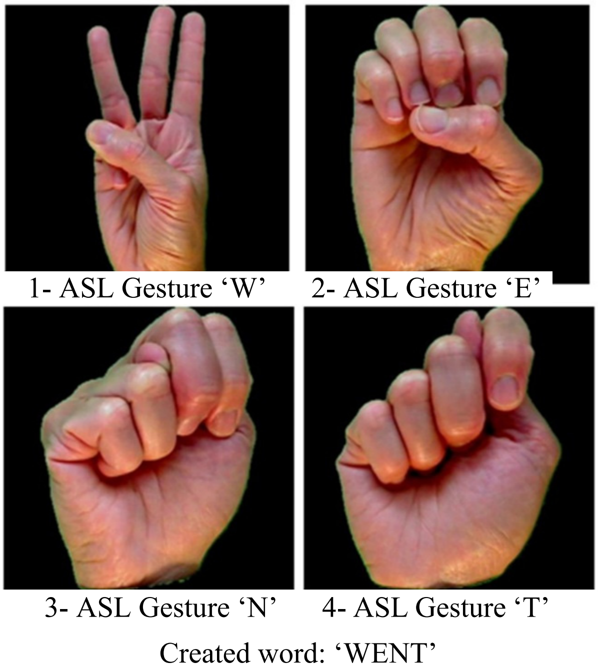


Figure 4.32. Creating word (ASL) scenario 1.

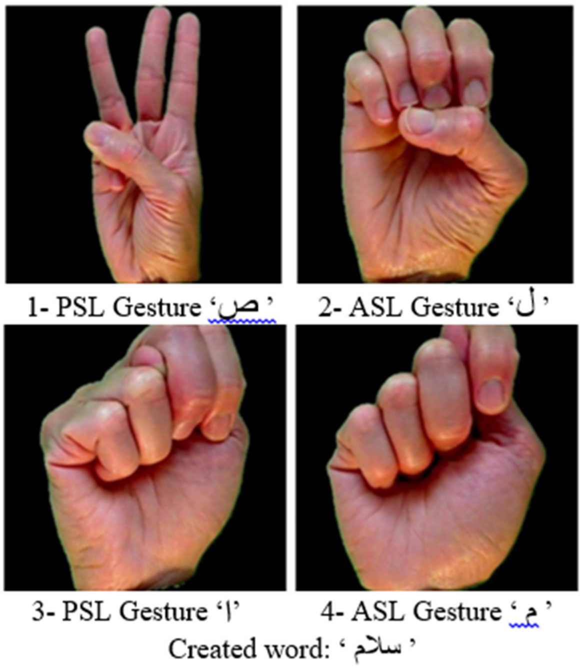


Figure 4.33. Creating word (PSL) scenario 2.

Figure 4.34 shows a word is created from gestures detected from SSL dataset. The word is pronounced as ‘Calle’ in Spanish. This is the translated in English as ‘Street’.


1- SSL Gesture ‘Ce’ 2- SSL Gesture ‘A’


3- SSL Gesture ‘Ele’ (twice) 4- SSL Gesture ‘E’

Created word: ‘Calle’

Figure 4.34. Creating word (SSL) scenario 3.

4.4. Algorithm Processing

In this section the different steps used for processing the hand gesture are discussed graphically. The unique feature if this system is that it can automatically detect multiple sign language datasets. The algorithm takes up to 5 gestures to decide which sign language datasetis used. Once the sign language dataset is selected, the algorithm process the acquired image using the dataset. In case the sign language dataset is detected after processing multiple gestures i.e. 2 to 5 then the algorithm goes back and reprocess the previous gestates (up tolast 5) to check if there is an error that can be corrected. For this purpose, one hand gesture i.e. ‘number 1’ from ASL dataset is used. This section presents how the acquired image is processed and information is extracted. The input image size is 400x400 pixels.

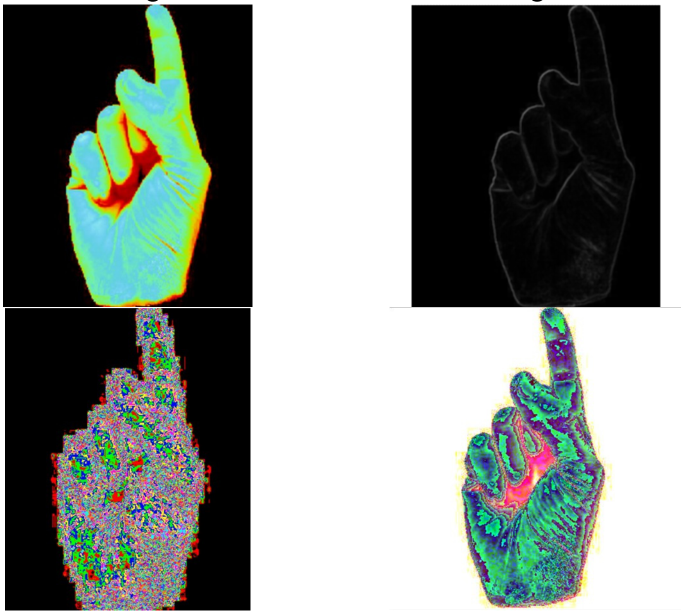
The acquired hand gesture image goes through processing before CNN algorithm using ReLU activation function. Figure 4.35 shows the acquired hand gesture image. This image is first converted into Grayscale which is then used for extracting some features from the image.


1. Acquired hand gesture Image 2. Grayscale

Figure 4.35. Original image and conversion to grayscale.

In the next step the image is further processed to extract more features including detection of edges and background removal. Removing background is an important feature which increases correct gesture detection accuracy. The images shown in figure 4.36 shows different processing stages which are focused on detecting gesture edges. The focus at this stage is to detect the position of the fingers. The last image shows the image with backgroundremoved.

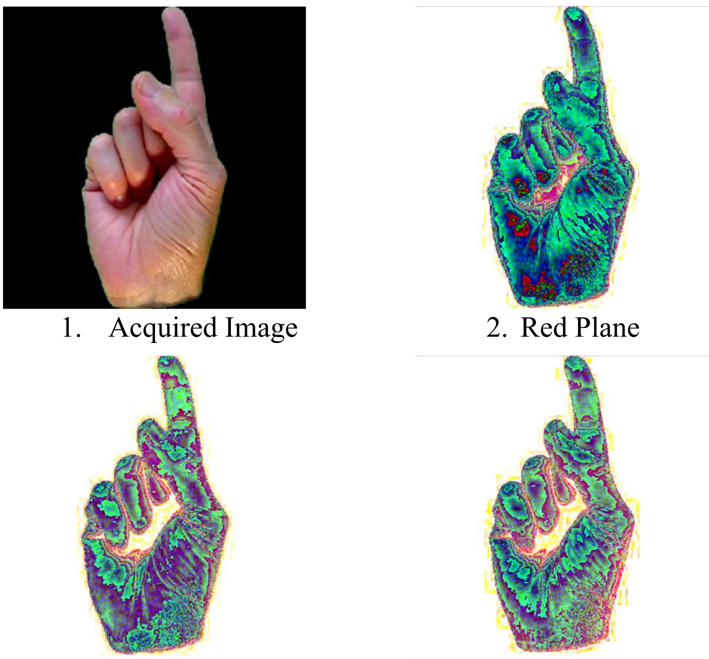
a. Extracting features # 1 b. Extracting features # 2



c. Extracting features # 3 d. Background removed

Figure 4.36. Features extracted from acquired hand gesture.

Figure 4.37 presents the original acquired hand gesture image and the RED, GREE and BLUE planes. The RGB planes are extracted after removing the background of the image.



1. Acquired Image 2. Red Plane

3. Green Plane 4. Blue Plane

Figure 4.37. Extract of RGB planes.

The next step is to apply CNN on the image using an activation function. The activation function used here is Rectified Linear Unit (ReLU) function. The ReLU function use a ‘0’ or a ‘1’ to process the input image. If the input image value is positive then the ReLU value is ‘1’ while a ‘0’ is used for zero or negative image value.

$$ReLU\ activation\ function = \left(\begin{matrix} 0\ (image\ value\ negative\ OR\ zero) \\ 1\ (image\ value\ positive) \end{matrix} \right) \quad (3)$$

Figure 4.38 shows the output image after applying CNN using ReLU activation function. This was done after the background is removed. It can be seen from the image that the edges are clearly visible.

Figure 4.38. CNN applied using ReLU.

The process of applying CNN on the acquired image is shown below. The image data is processed using a kernel as shown in figure 4.39 and equation (4). The final value isrepresented by ‘G’ in the equation.

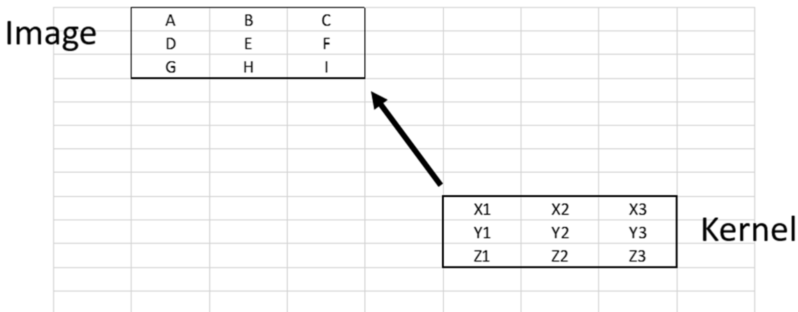


Figure 4.39. Applying CNN using ReLU.

The following equation shows the calculation for each value generated after processing.

$$G = (A * X1) + (B * X2) + (C * X3) + (D * Y1) + (E * Y2) + (F * Y3) + (G * Z1) + (H * Z2) + (I * Z3) \quad (4)$$

Figure 4.40. shows the portion of the input image where the algorithm is applied to show how the algorithm is processed.

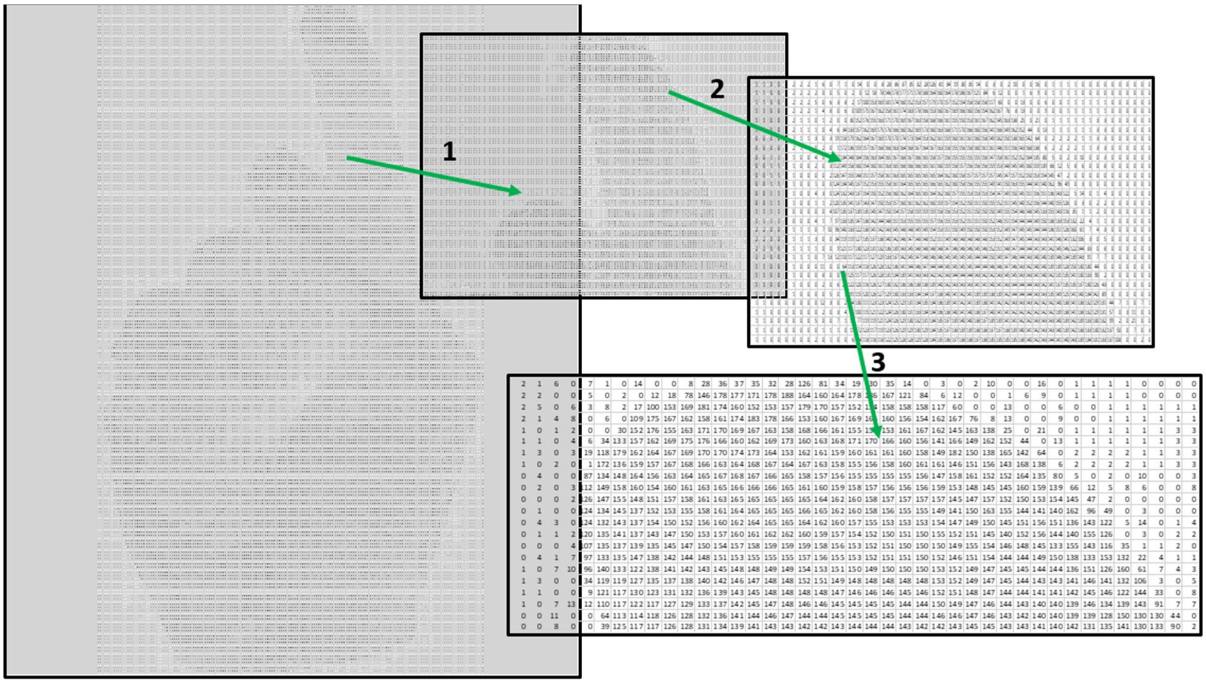
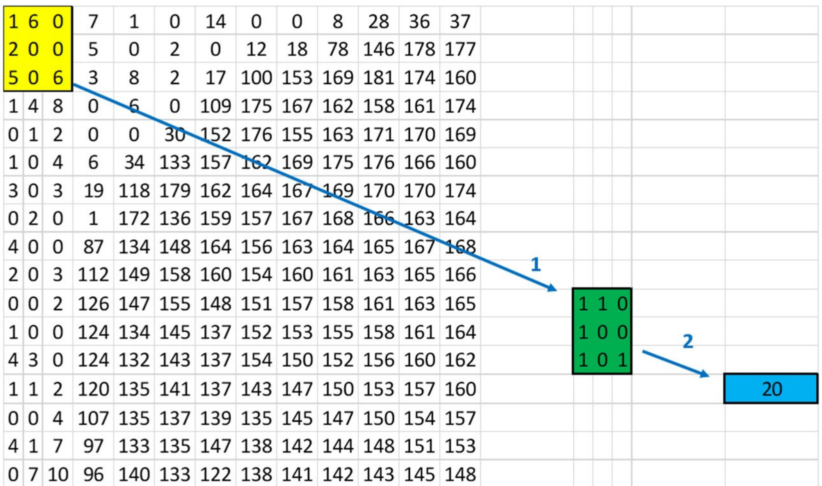


Figure 4.40. Input image used for algorithm processing.

Figures 4.41 till 4.52 shows the steps used in processing input image using the kernel. Each figure shows the portion of the image used for processing along with the values used for the kernel. The final calculated value is shown on the right bottom corner. Equation 5 shows the calculation for image 4.41.

$$G = (1 * 1) + (6 * 1) + (0 * 0) + (2 * 1) + (0 * 0) + (0 * 0) + (5 * 1) + (0 * 0) + (6 * 1) = 20 \quad (5)$$



1	6	0	7	1	0	14	0	0	8	28	36	37
2	0	0	5	0	2	0	12	18	78	146	178	177
5	0	6	3	8	2	17	100	153	169	181	174	160
1	4	8	0	6	0	109	175	167	162	158	161	174
0	1	2	0	0	30	152	176	155	163	171	170	169
1	0	4	6	34	133	157	162	169	175	176	166	160
3	0	3	19	118	179	162	164	167	169	170	170	174
0	2	0	1	172	136	159	157	167	168	166	163	164
4	0	0	87	134	148	164	156	163	164	165	167	168
2	0	3	112	149	158	160	154	160	161	163	165	166
0	0	2	126	147	155	148	151	157	158	161	163	165
1	0	0	124	134	145	137	152	153	155	158	161	164
4	3	0	124	132	143	137	154	150	152	156	160	162
1	1	2	120	135	141	137	143	147	150	153	157	160
0	0	4	107	135	137	139	135	145	147	150	154	157
4	1	7	97	133	135	147	138	142	144	148	151	153
0	7	10	96	140	133	122	138	141	142	143	145	148

Figure 4.41. Graphical representation step 1.

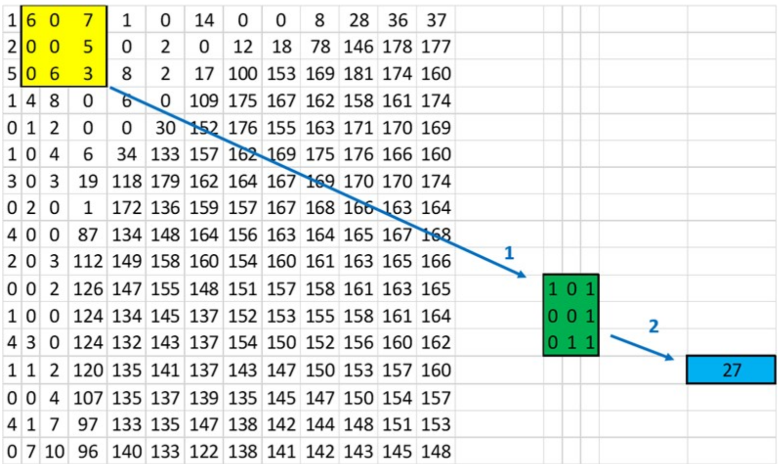


Figure 4.42. Graphical representation step 2.

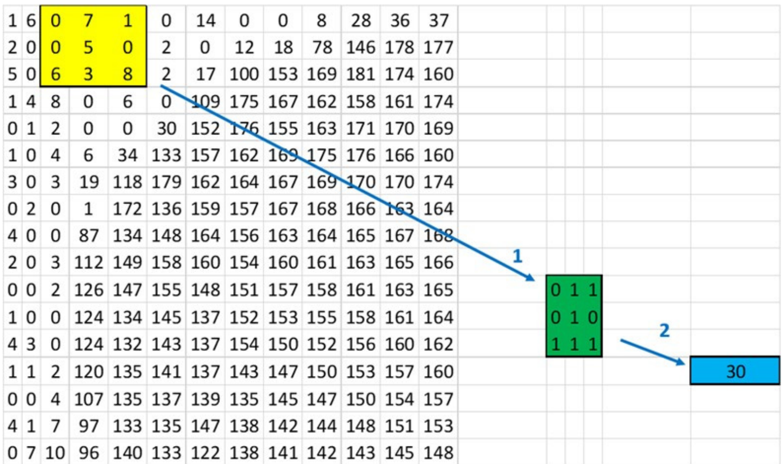


Figure 4.43. Graphical representation step 3.

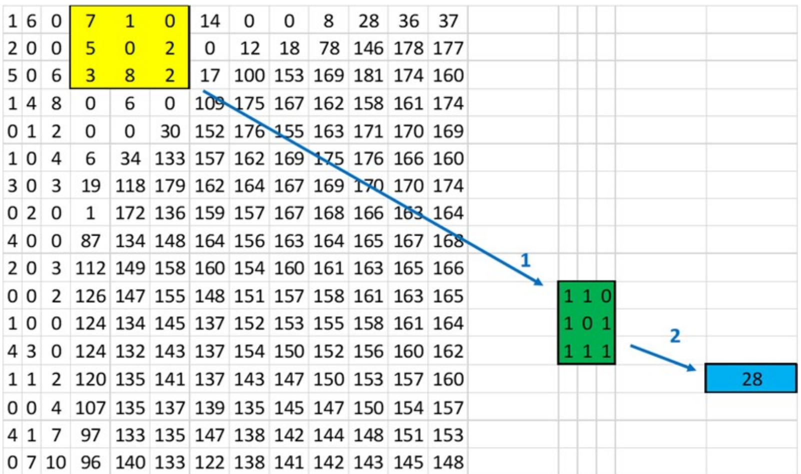


Figure 4.44. Graphical representation step 4.

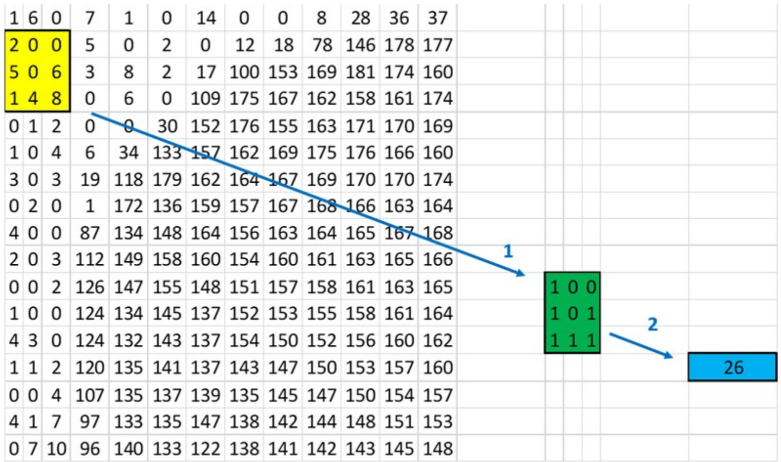


Figure 4.45. Graphical representation step 5.

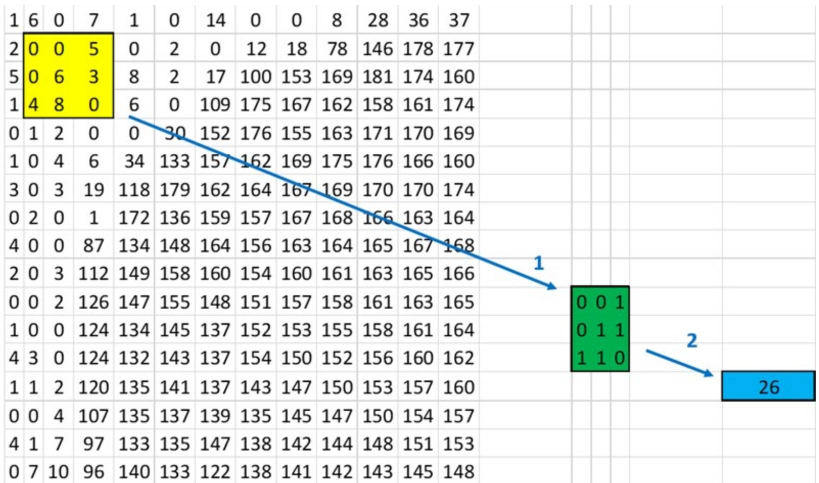


Figure 4.45. Graphical representation step 6.

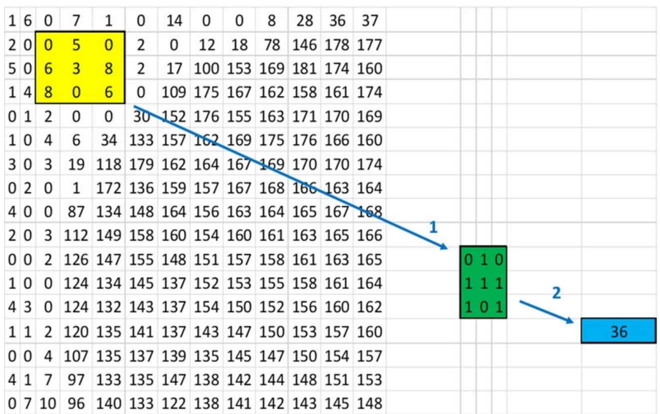


Figure 4.45. Graphical representation step 7.

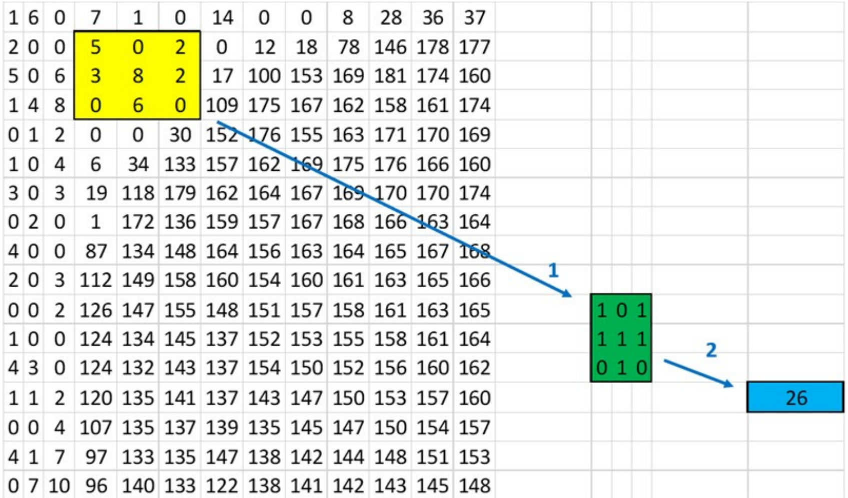


Figure 4.48. Graphical representation step 8.

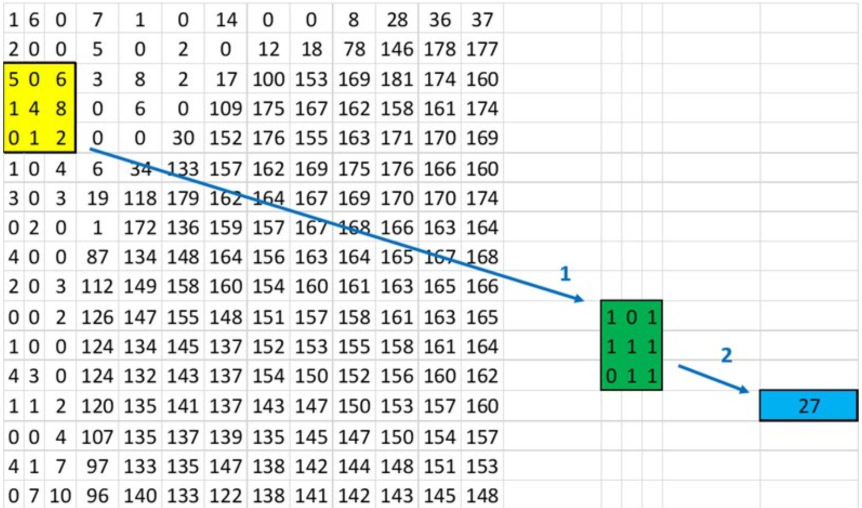


Figure 4.49. Graphical representation step 9.

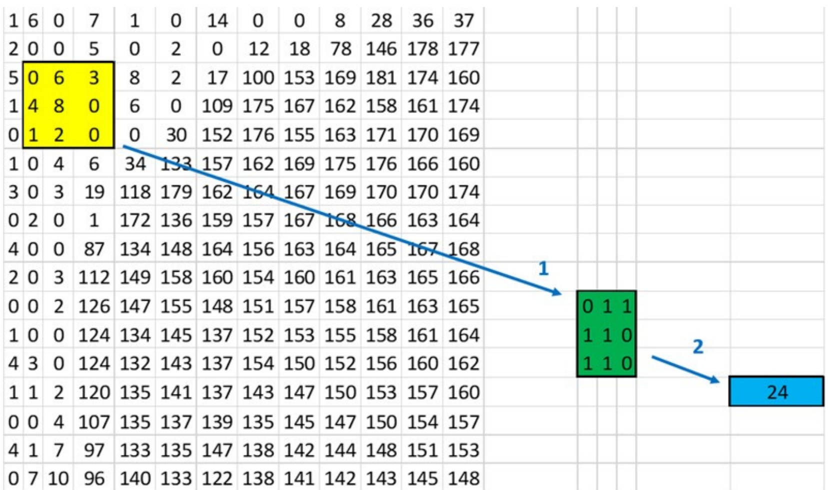


Figure 4.50. Graphical representation step 10.

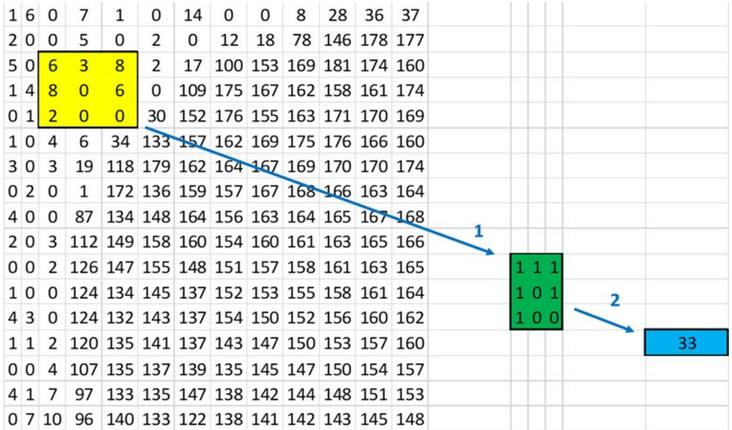


Figure 4.51. Graphical representation step 11.

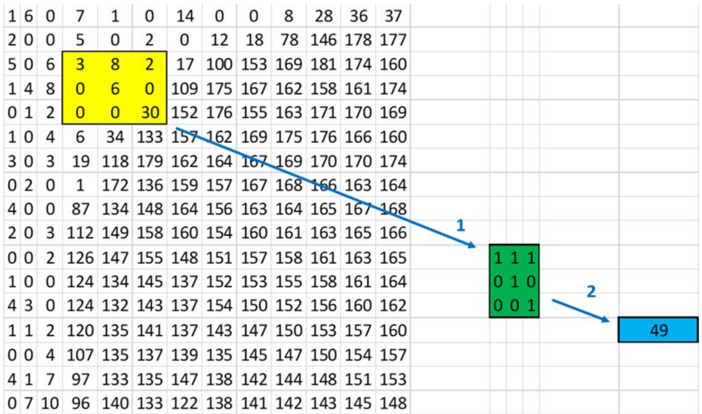


Figure 4.52. Graphical representation step 12.

The final feature map is shown in the figure 4.53. This feature map is for the portion of the image processed.

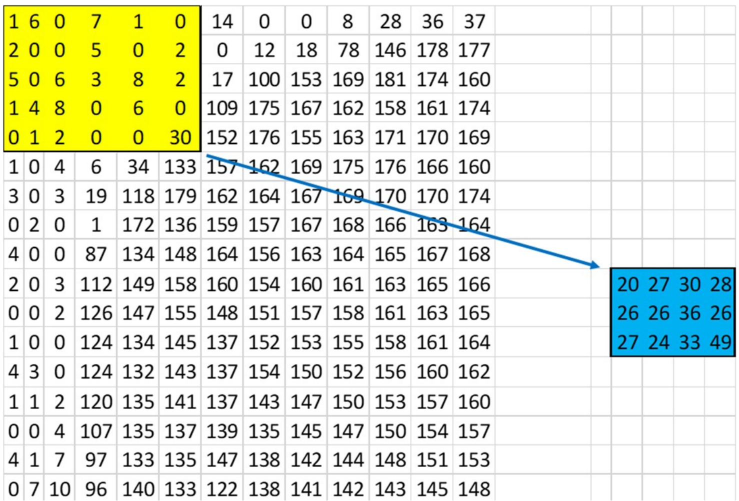


Figure 4.53. Final calculated values for the processed portion of image.

4.4.1. Edge detection

In this section the results of edges detected for various gestures are presented. This is an important feature required for detecting the position of fingers and thumb. The edges extracted and the output of CNN are both shown in the figures 4.54 till 4.65.

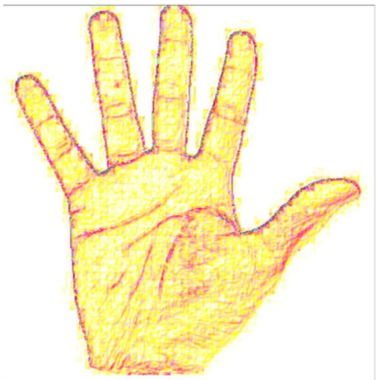


Figure 4.54. CNN output for ASL number '5'.

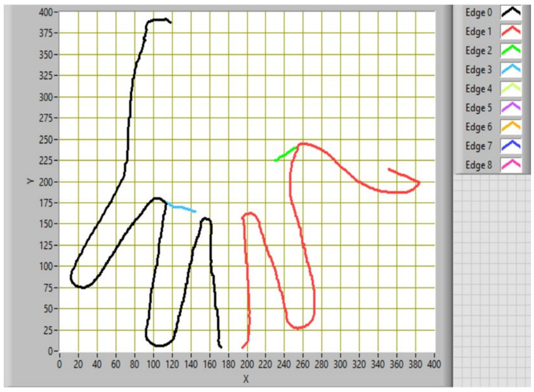


Figure 4.55. Edges detected for ASL number '5'.

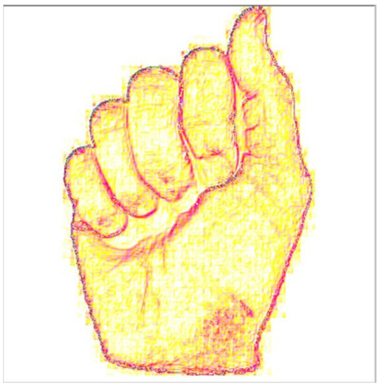


Figure 4.56. CNN output for ASL letter ‘A’.

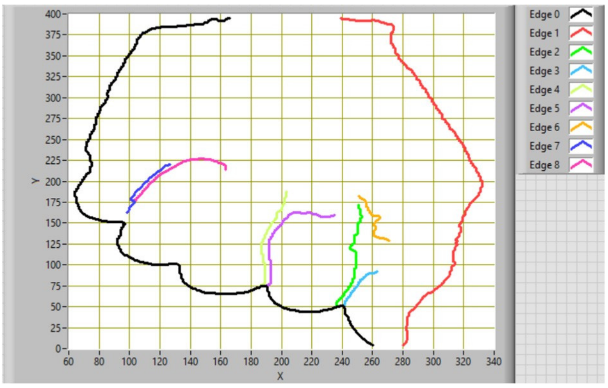


Figure 4.57. Edges detected for ASL letter ‘A’.

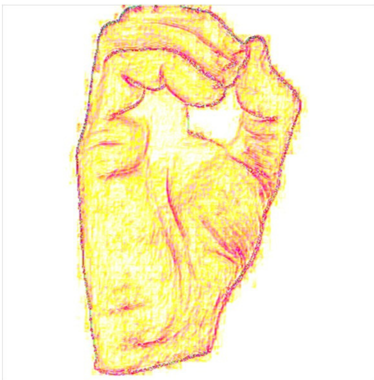


Figure 4.58. CNN output for ASL number ‘0’.

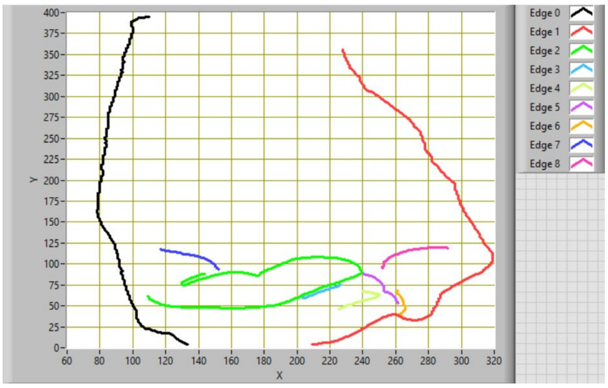


Figure 4.59. Edges detected for ASL number ‘0’.

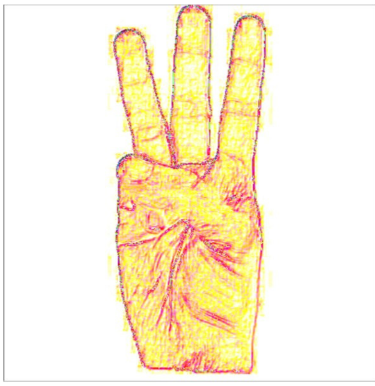


Figure 4.60. CNN output for ASL letter ‘W’.

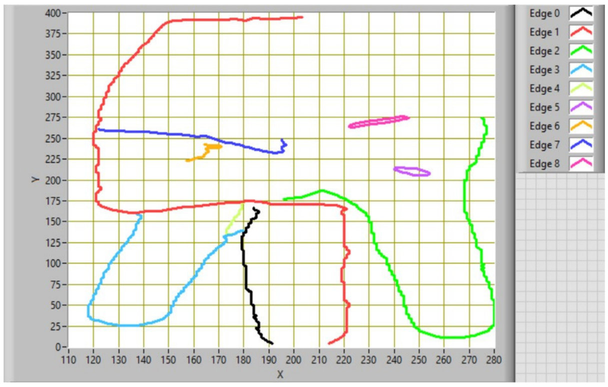


Figure 4.61. Edges detected for ASL letter ‘W’.

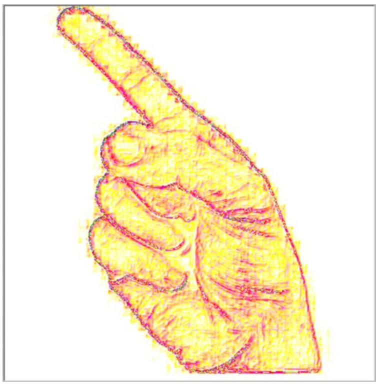


Figure 4.62. CNN output for ASL letter ‘Z’.

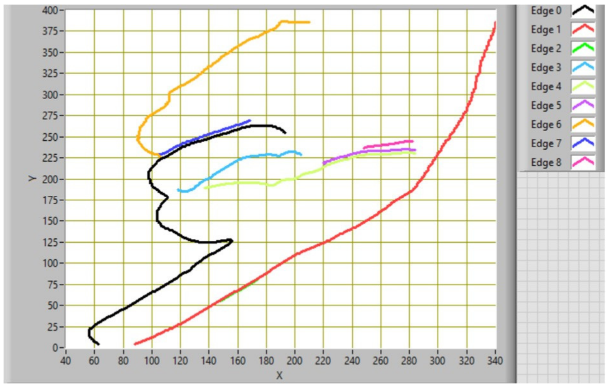


Figure 4.63. Edges detected for ASL letter ‘Z’.

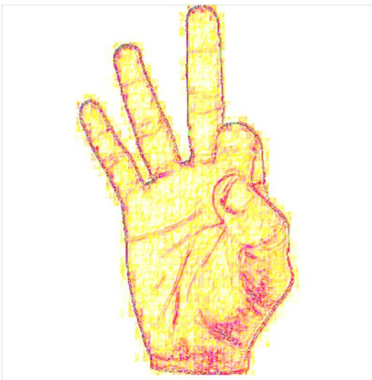


Figure 4.64. CNN output for ASL number ‘9’.

Figure 4.65. Edges detected for ASL number ‘9’

Likewise, edge detection is employed on both PSL and ASL datasets. The work encompasses various aspects such as machine learning (ML), convolutional neural networks (CNN), the utilization of multiple Sign Language datasets, and commercial off-the-shelf (COTS) hardware devices. The system described in this manuscript builds upon prior research that has been published. The system utilizes a combination of three sign language datasets. The system has a distinctive capability to automatically identify and interpret sign language. The related work section includes a thorough analysis of the existing systems, which is presented in an extensive review. The review is accompanied by a tabular summary that provides a concise overview of the analysis and highlights key features of the existing systems. The user's text encompasses an examination of various datasets pertaining to different sign languages, accompanied by an assessment of machine learning methods.

Chapter 5: Conclusions and Future Research

5.1. Conclusions

D-M people are an integral part of society, and the system presented in this manuscript aims to help them communicate with ND-M people.

(a) Summary of the key findings and results obtained during the research:

In this thesis, a computer-based systemprovides two-way communication between D-M people and ND-M people. The D-M interface is provided so that their hand gestures are acquired and converted into text and then voice. The ND-M people are provided with an audio interface in which their voices are acquired and then converted into text. The system uses deep learning to detect and process hand gestures, which are implemented using a Leap device. Language detection is achieved through supervised machine learning. The system also provides speech-to-text-to-speechconversion, which is a useful feature for ND-M people. Due to this option, ND-M people do not need any training or are not required to know sign language.

(b) An implications and significance of the research in the broader context of the field of Study:
The feature of the proposed system is that the end user is not expected to undergo rigorous training. Ease of use is a key feature of this research. Another key feature is that D- M individuals can also use this system to communicate with each other. In this learningmode, the system will acquire the data, which will help improve the performance of the machine learning algorithm. The software application is easy to install and doesn't requireany licensing. The software processes new data and stores them in the database.

The scope of this work includes ML, CNN, the use of multiple Sign Language datasets, and COTS hardware devices. The system presented in this manuscript is the continuation of previously published work. The system uses three sign language datasets combined into one. This is a unique feature where the system automatically detects sign language. An extensive review of the existing systems is carried out and discussed in the related work section, including a tabular summary providing a summary of the review and listing some of the features of the existing systems. This also includes an analysis of some available datasets of different sign languages, followed by an evaluation of some machine learning techniques.

The proposed system is developed, implemented, and then validated using a four-stageresearch methodology. The key features of the proposed system include the implementationof CNN, ML, text-to-speech-to-text, a training interface, and a database storing the sign language datasets and user profiles.

Related to the design phase, the system acquires hand gesture data using a COTS LMD device (Leap motion), and then the data is processed using the CNN algorithm. Using a COTS device means that there is no need for extensive validation and verification to convert a prototype into a product, because has a stable hardware interface. The other advantage is that there is no calibration required compared to other custom-made devices, which are normally based on sensors connected to a glove and require continuous maintenance to make sure the sensors are working correctly. Upgradation of custom-made devices is also a challenge where the availability of components has been a problem recently. The software is developed in LabVIEW, which provides a high-quality graphical user interface (GUI) and quick development.

The system is validated through a series of experiments, which include hand gesture detection accuracy, processing of individual hand gesture data, evaluation of hand gestures with varying image quality, detection of variations in the same gestures, and the identification and processing of different hand gestures that look similar, detection of hand gestures with visibleobjects such as a watch, ring, etc., and finally automatic identification of a sign language dataset. The limitations of some features are also defined. Additionally, due to the system stores the acquired data and uses it for processing, the system accuracy will increase due to more data being added to the dataset.

Another key point is the performance, which has been evaluated using several experiments, namely detection of similar hand gestures, low-resolution gesture images, inaccurate gestures, and speed of making hand gestures. As general result, the system can detect hand gestures correctly, and the overall accuracy is mostly more than 90%. There are a few scenarios in which the accuracy is between 80% and 90% for similar gestures. The mathematical tool used was the confusion matrix.

(c) A final remarks and conclusion of the thesis and its overall contribution:
As final remark, many existing similar systems only focus on D-M people, where the interface is provided for them so that their hand gestures are detected. These systems expect ND-M people to learn about the detection of hand gestures and the interfaces provided to them. It is not easy for ND-M people to learn sign language, hence reducing the effectiveness of this system. The proposed system provides a complete solution to the problems faced by both D-M and ND-M people.

5.2. Future Research

One of the novel features is that the system supports multiple languages. In the future, the GUI can be modified to add more language gestures to the database to support a multi- language communication. The challenge will be to process similar gestures from different languages.

In the scope of the AI technology, the dataset size can be increased based on the recommendations from ML-based algorithms. More work can be done in the undertaking of comparisons between different sign languages, understanding similarities between them, and afterward combining these to create larger datasets.

Another interesting capability to be considered, is to use a new dataset improved by adding videos and other types of data, including word-level hand gestures. That enhanced database could extend the experience of communication between D-M and ND-M people towards a more natural interaction. In this sense, the customization could be a desired feature for specific D-M people, for example considering extra time to acquire every gesture due to a limited hand movement as consequence of injury or disability.

In the option of training features, there are more options to consider in a near future. For example, a child/beginner training must have different requirements for the system that an experimented D-M person, who already knows the sign language.

Bibliography

[1] Vaidya, O.; Gandhe, S.; Sharma, A.; Bhate, A.; Bhosale, V.; Mahale, R. Design and Development of Hand Gesture based Communication Device for Deaf and Mute People. In Proceedings of the IEEE Bombay Section Signature Conference (IBSSC), Mumbai, India, 4–6 December 2020; pp. 102–106. https://doi.org/10.1109/IBSSC51096.2020.9332208.

[2] Marin, G.; Dominio, F.; Zanuttigh, P. Hand gesture recognition with Leap Motion and Kinect devices. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 1565–1569. https://doi.org/10.1109/ICIP.2014.7025313.

[3] Nwankpa, C. E., Ijomah, W., Gachagan, A., & Marshall, S. (2021). Activation functions: Comparison of trends in practice and research for deep learning. 124 - 133. 2nd International Conference on Computational Sciences and Technology. https://doi.org/10.48550/arXiv.1811.03378

[4] Saleem, M.I.; Siddiqui, A.A.; Noor, S.; Luque-Nieto, M.Á. A Novel Machine Learning Based Two-Way Communication System for Deaf and Mute. *Appl. Sci.* 2022, *13*, 453. https://doi.org/10.3390/app13010453.

[5] Deb, S.; Suraksha; Bhattacharya, P. Augmented Sign Language Modeling (ASLM) with interaction design on smartphone—An assistive learning and communication tool for inclusive classroom. *Procedia Comput. Sci.* 2018, *125*, 492–500. ISSN 1877-0509. https://doi.org/10.1016/j.procs.2017.12.064.

[6] Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. Computer Science—Computer Vision and Pattern Recognition. *arXiv* 2017, arXiv:1704.04861.

[7] Rishi, K.; Prarthana, A.; Pravena, K.S.; Sasikala, S.; Arunkumar, S. Two-Way Sign Language Conversion for Assisting Deaf-Mutes Using Neural Network. In Proceedings of the 8th International Conference on Advanced Computing and Communication Systems, ICACCS 2022, Coimbatore, India, 25–26 March 2022; pp. 642–646. https://doi.org/10.1109/ICACCS54159.2022.9785057.

[8] Anupama, H.S.; Usha, B.A.; Madhushankar, S.; Vivek, V.; Kulkarni, Y. Automated Sign Language Interpreter Using Data Gloves. In Proceedings of the International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 25–27 March 2021; pp. 472–476. https://doi.org/10.1109/ICAIS50930.2021.9395749.

[9] Kawai, H.; Tamura, S. Deaf-and-mute sign language generation system. *Pattern Recognit.* 1985, *18*, 199–205. https://doi.org/10.1016/0031-3203(85)90045-7.

[10] Bhadauria, R.S.; Nair, S.; Pal, D.K. A Survey of Deaf Mutes. *Med. J. Armed Forces India* 2007, *63*, 29–32. https://doi.org/10.1016/S0377-1237(07)80102-X.

[11] Sood, A.; Mishra, A. AAWAAZ: A communication system for deaf and dumb. In Proceedings of the 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 7–9 September 2016; pp. 620–624. https://doi.org/10.1109/ICRITO.2016.7785029.

[12] Yousaf, K.; Mehmood, Z.; Saba, T.; Rehman, R.; Rashid, M.; Altaf, M.; Shuguang, Z. A Novel Technique for Speech Recognition and Visualization Based Mobile Application to Support Two-Way Communication between Deaf-Mute and Normal Peoples. *Wirel.Commun. Mob. Comput.* 2018, *2018*, 1013234. https://doi.org/10.1155/2018/1013234.

[13] Raheja, J.L.; Singhal, A.; Chaudhary, A. Android Based Portable Hand Sign Recognition System; Machine Vision Lab, CSIR-CEERI, Pilani, India. *arXiv* 2015, arXiv:1503.03614.

[14] Soni, N.S.; Nagmode, M.S.; Komati, R.D. Online hand gesture recognition & classification for deaf & dumb. In Proceedings of the International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 26–27 August 2016; pp. 1–4. https://doi.org/10.1109/INVENTIVE.2016.7830112.

[15] Chakrabarti, S. State of deaf children in West Bengal, India: What can be done to improve outcome. *Int. J. Pediatr. Otorhinolaryngol.* 2018, *110*, 3742. https://doi.org/10.1016/j.ijporl.2018.04.025.

[16] Ameur, S.; Khalifa, A.B.; Bouhlel, M.S. Chronological pattern indexing: An efficient feature extraction method for hand gesture recognition with Leap Motion. *J. Vis. Commun. Image Represent.* 2020, *70*, 102842. https://doi.org/10.1016/j.jvcir.2020.102842.

[17] Ameur, S.; Khalifa, A.B.; Bouhlel, M.S. A novel hybrid bidirectional unidirectional LSTM network for dynamic hand gesture recognition with Leap Motion. *Entertain. Comput.* 2020, *35*, 100373. https://doi.org/10.1016/j.entcom.2020.100373.

[18] Boppana, L.; Ahamed, R.; Rane, H.; Kodali, R.K. Assistive Sign Language Converter for Deaf and Dumb. In Proceedings of the 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Atlanta, GA, USA, 14–17 July 2019; pp. 302–307. https://doi.org/10.1109/iThings/GreenCom/CPSCom/SmartData.2019.00071.

[19] Suharjito; Anderson, R.; Wiryana, F.; Ariesta, M.C.; Kusuma, G.P. Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process- Output. *Procedia Comput. Sci.* 2017, *116*, 441–448. ISSN 1877-0509. https://doi.org/10.1016/j.procs.2017.10.028.

[20] Patwary, A.S.; Zaohar, Z.; Sornaly, A.A.; Khan, R. Speaking System for Deaf and Mute People with Flex Sensors. In Proceedings of the 2022 6th International Conference on Trends in Electronics and Informatics, ICOEI 2022-Proceedings; Tirunelveli, India, 28–30 April 2022; pp. 168–173. https://doi.org/10.1109/ICOEI53556.2022.9777226.

[21] Sharma, A.; Yadav, A.; Srivastava, S.; Gupta, R. Analysis of movement and gesture recognition using Leap Motion Controller. *Procedia Comput. Sci.* 2018, *132*, 551–556. ISSN 1877-0509. https://doi.org/10.1016/j.procs.2018.05.008.

[22] Salem, N.; Alharbi, S.; Khezendar, R.; Alshami, H. Real-time glove and android application for visual and audible Arabic sign language translation. *Procedia Comput. Sci.* 2019, *163*, 450–459. ISSN 1877-0509. https://doi.org/10.1016/j.procs.2019.12.128.

[23] Zhang, Y.; Min, Y.; Chen, X. Teaching Chinese Sign Language with a Smartphone. *Virtual Real. Intell. Hardw.* 2021, *3*, 248–260. ISSN 2096-5796. https://doi.org/10.1016/j.vrih.2021.05.004.

[24] Samonte, M.J.C.; Gazmin, R.A.; Soriano, J.D.S.; Valencia, M.N.O. BridgeApp: An Assistive Mobile Communication Application for the Deaf and Mute. In Proceedings of the 2019 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea, 16–18 October 2019; pp. 1310–1315. https://doi.org/10.1109/ICTC46691.2019.8939866.

[25] Sobhan, M.; Chowdhury, M.Z.; Ahsan, I.; Mahmud, H.; Hasan, M.K. A Communication Aid System for Deaf and Mute using Vibrotactile and Visual Feedback. In Proceedings of the 2019 International Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, Indonesia, 21–22 September 2019; pp. 184–190. https://doi.org/10.1109/ISEMANTIC.2019.8884323.

[26] Sunil Kumar, K.N.; Sathish, R.; Vinayak, S.; Pandit, T.P. Braille Assistance System for Visually Impaired, Blind & Deaf-Mute people in Indoor & Outdoor Application. In Proceedings of the 2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), Bangalore, India, 17–18 May 2019; pp. 1505–1509. https://doi.org/10.1109/RTEICT46194.2019.9016765.

[27] Villagomez, E.B.; King, R.A.; Ordinario, M.J.; Lazaro, J.; Villaverde, J.F. Hand Gesture Recognition for Deaf-Mute using FuzzyNeural Network. In Proceedings of the 2019 IEEE International Conference on Consumer Electronics—Asia (ICCE-Asia), Bangkok, Thailand, 12–14 June 2019; pp. 30–33. https://doi.org/10.1109/ICCEAsia46551.2019.8942220.

[28] Tao, Y.; Huo, S.; Zhou, W. Research on Communication APP for Deaf and Mute People Based on Face Emotion Recognition Technology. In Proceedings of the 2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Weihai, China, 14–16 October 2020; pp. 547–552. https://doi.org/10.1109/ICCASIT50869.2020.9368771.

[29] Shareef, S.K.; Haritha, I.V.S.L.; Prasanna, Y.L.; Kumar, G.K. Deep Learning Based Hand Gesture Translation System. In Proceedings of the 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 3–5 June 2021; pp. 1531– 1534. https://doi.org/10.1109/ICOEI51242.2021.9452947.

[30] Dhruv, A.J.; Bharti, S.K. Real-Time Sign Language Converter for Mute and Deaf People. In Proceedings of the 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), Gandhinagar, India, 24–26 September 2021; pp. 1–6. https://doi.org/10.1109/AIMV53313.2021.9670928.

[31] Rosero-Montalvo, P.D.; Godoy-Trujillo, P.; Flores-Bosmediano, E.; Carrascal-Garcia, J.; Otero-Potosi, S.; Benitez-Pereira, H.; Peluffo-Ordonez, D.H. Sign Language RecognitionBased on Intelligent Glove Using Machine Learning Techniques. In Proceedings of the 2018 IEEE Third Ecuador Technical Chapters Meeting (ETCM), Cuenca, Ecuador, 15–19 October 2018; pp. 1–5. https://doi.org/10.1109/ETCM.2018.8580268.

[32] Janeera, D.A.; Raja, K.M.; Pravin, U.K.R.; Kumar, M.K. Neural Network based Real Time Sign Language Interpreter for Virtual Meet. In Proceedings of the 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 8–10 April 2021; pp. 1593–1597. https://doi.org/10.1109/ICCMC51019.2021.9418238.

[33] Gupta, A.M.; Koltharkar, S.S.; Patel, H.D.; Naik, S. DRISHYAM: An Interpreter for Deaf and Mute Using Single Shot Detector Model. In Proceedings of the 8th International Conference on Advanced Computing and Communication Systems, ICACCS 2022, Coimbatore, India, 25–26 March 2022; pp. 365–371. https://doi.org/10.1109/ICACCS54159.2022.9785254.

[34] Lan, S.; Ye, L.; Zhang, K. Attention-Augmented Electromagnetic Representation of Sign Language for Human-Computer Interaction in Deafand-Mute Community. In Proceedings of the 2021 IEEE USNC-URSI Radio Science Meeting (Joint with AP-S Symposium), Singapore, Singapore, 4–10 December 2021; pp. 47–48. https://doi.org/10.23919/USNCURSI51813.2021.9703456.

[35] Telluri, P.; Manam, S.; Somarouthu, S.; Oli, J.M.; Ramesh, C. Low cost flex powered gesture detection system and its applications. In Proceedings of the 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 15–17 July 2020; pp. 1128–1131. https://doi.org/10.1109/ICIRCA48905.2020.9182833.

[36] Jamdar, V.; Garje, Y.; Khedekar, T.; Waghmare, S.; Dhore, M.L. Inner Voice—An Effortless Way of Communication For The Physically Challenged Deaf & Mute People. In Proceedings of the 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), Gandhinagar, India, 24–26 September 2021; pp. 1–5. https://doi.org/10.1109/AIMV53313.2021.9670911.

[37] He, Y.; Kuerban, A.; Yu, Q.; Xie, Q. Design and Implementation of a Sign Language Translation System for Deaf People. In Proceedings of the 2021 3rd International Conference on Natural Language Processing (ICNLP), Beijing, China, 26–28 March 2021; pp. 150–154. https://doi.org/10.1109/ICNLP52887.2021.00031.

[38] Xia, K.; Lu, W.; Fan, H.; Zhao, Q.A. Sign Language Recognition System Applied to Deaf-Mute Medical Consultation. *Sensors* 2022, *22*, 9107. https://doi.org/10.3390/s22239107.

[39] Siddiqui, A.; Zia, M.Y.I.; Otero, P. A universal machine-learning-based automated testing system for consumer electronic products. *Electronics* 2021, *10*, 136. https://doi.org/10.3390/electronics10020136.

[40] Siddiqui, A.; Zia, M.Y.I.; Otero, P. A Novel Process to Setup Electronic Products Test Sites Based on Figure of Merit and Machine Learning. *IEEE Access* 2021, *9*, 80582–80602. https://doi.org/10.1109/ACCESS.2021.3084545.

[41] Ronchetti, F.; Quiroga, F.; Estrebou, C.A.; Lanzarini, L.C.; Rosete, A. LSA64: An Argentinian Sign Language Dataset. *Congreso Argentino de Ciencias de La Computacion (CACIC)*, San Luis, Argentina, 3–10 October 2016; pp. 794–803.

[42] Joze, H.R.V.; Koller, O. MS-ASL: A Large-Scale Data Set and Benchmark for Understanding American Sign Language. In Proceedings of the 30th British Machine Vision Conference 2019, BMVC 2019, Cardiff, UK, 9–12 September 2019.

[43] Jie, H.; Zhou, W.; Li, H.; Li, W. Attention-Based 3D-CNNs for Large-Vocabulary Sign Language Recognition. *IEEE Trans. Circuits Syst. Video Technol.* 2019, *29*, 2822–2832. https://doi.org/10.1109/TCSVT.2018.2870740.

[44] Kagirov, I.; Ivanko, D.; Ryumin, D.; Axyonov, A.; Karpov, A. TheRuSLan: Database of Russian Sign Language. In Proceedings of the LREC 2020—12th International Conference on Language Resources and Evaluation, Conference Proceedings, Marseille, France, 11–16 May 2020; pp. 6079–6085.

[45] Sincan, O.M.; Keles, H.Y. AUTSL: A Large Scale Multi-Modal Turkish Sign Language Dataset and Baseline Methods. *IEEE Access* 2020, *8*, 181340–181355. https://doi.org/10.1109/ACCESS.2020.3028072.

[46] Dongxu, L.; Opazo, C.R.; Yu, X.; Li, H. Word-Level Deep Sign Language Recognition from Video: A New Large-Scale Dataset and Methods Comparison. 2020. Available online: https://dxli94.github.io/ (accessed on 28 December 2022).

[47] Tavella, F.; Schlegel, V.; Romeo, M.; Galata, A.; Cangelosi, A. WLASL-LEX: A Dataset for Recognising Phonological Properties in American Sign Language. *arXiv* 2022; pp. 453–463, https://doi.org/10.18653/v1/2022.acl-short.49.

[48] Saleem, M.I.; Otero, P.; Noor, S.; Aftab, R. Full Duplex Smart System for Deaf Dumb and Normal People. In Proceedings of the 2020 Global Conference on Wireless and Optical Technologies, GCWOT 2020. 2020. https://doi.org/10.1109/GCWOT49901.2020.9391593.

[49] Asthana, O.; Bhakuni, P.; Srivastava, P.; Singh, S.; Jindal, K. Sign Language Recognition Based on Gesture Recognition/Holistic Features Recognition: A Review of Techniques. In Proceedings of the 2nd International Conference on Innovative Practices in Technology and Management, ICIPTM 2022, 2022; pp. 713–718. https://doi.org/10.1109/ICIPTM54933.2022.9754140.

[50] Farhan, Y.; Madi, A.A.; Ryahi, A.; Derwich, F. American Sign Language: Detection and Automatic Text Generation. In Proceedings of the 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology, IRASET 2022, 2022; pp. 1–6. https://doi.org/10.1109/IRASET52964.2022.9738061.

[51] Bisht, D.; et al. Smart Communication System Using Sign Language Interpretation. In Proceedings of the Conference of Open Innovation Association, FRUCT. FRUCT Oy, April 2022; pp. 12–20. https://doi.org/10.23919/FRUCT54823.2022.9770914.

[52] Billah, A.R.; Fahad, F.I.; Raaz, S.R.; Saha, A.; Eity, Q.N. Recognition of Bangla SignLanguage Characters and Digits Using CNN. In Proceedings of the 2022 International Conference on Innovations in Science, Engineering and Technology, ICISET 2022, February 2022; pp. 239–244. https://doi.org/10.1109/ICISET54810.2022.9775823.

[53] Ben Hamouda, S.; Gabsi, W. Arabic sign Language Recognition: Towards a Dual Way Communication System between Deaf and Non-Deaf People. In Proceedings of the 22nd IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2021-Fall, 2021; Institute of Electrical and Electronics Engineers Inc.: 2021; pp. 37–42. https://doi.org/10.1109/SNPD51163.2021.9705002.

[54] Illahi, A.A.C.; Betito, M.F.M.; Chen, C.C.F.; Navarro, C.V.A.; Or, I.V.L. Development of a Sign Language Glove Translator Using Microcontroller and Android Technology for Deaf-Mute. In Proceedings of the 2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management, HNICEM 2021, 2021; pp. 1–5. https://doi.org/10.1109/HNICEM54116.2021.9731835.

[55] Snehaa, A.; Suryaprakash, S.; Sandeep, A.; Monikapriya, S.; Mathankumar, M.; Thirumoorthi, P. Smart Audio Glove for Deaf and Dumb Impaired. In Proceedings of the2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation, ICAECA 2021, 2021. https://doi.org/10.1109/ICAECA52838.2021.9675729.

[56] Vanaja, S.; Preetha, R.; Sudha, S. Hand Gesture Recognition for Deaf and Dumb Using CNN Technique. In Proceedings of the 6th International Conference on Communication and Electronics Systems, ICCES 202, 2021. https://doi.org/10.1109/ICCES51350.2021.9489209.

[57] Ameur, S.; Khalifa, A.B.; and Bouhlel, M.S. "A comprehensive leap motion database for hand gesture recognition," *2016 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)*, Hammamet, Tunisia, 2016, pp. 514-519, doi: 10.1109/SETIT.2016.7939924.

[58] Morales, R.A.; Figueroa, M.A.A.; Mazahua, L.R.; Contreras, H.M.; and Martínez, U.J. "Architecture for the integration of the Leap Motion device into applicationdevelopment," *2021 10th International Conference On Software Process Improvement (CIMPS)*, Torreón, Coahuila, Mexico, 2021, pp. 1-9, doi: 10.1109/CIMPS54606.2021.9652726.

[59] Nosirov, K.; Begmatov, S.; and Arabboev, M. "Analog Sensing and Leap Motion Integrated Remote Controller for Search and Rescue Robot System," *2020 International Conference on Information Science and Communications Technologies (ICISCT)*, Tashkent, Uzbekistan,2020, pp. 1-5, doi: 10.1109/ICISCT50599.2020.9351425.

[60] Sharma, J.K.; Gupta, R.; and Pathak, V.K. "Numeral Gesture Recognition Using Leap Motion Sensor," *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, Jabalpur, India, 2015, pp. 411-414, doi: 10.1109/CICN.2015.86.

[61] Erdoğan, K.; Durdu, A.; and Yilmaz, N. "Intention recognition using leap motion controller and Artificial Neural Networks," *2016 International Conference on Control, Decision and Information Technologies (CoDIT)*, Saint Julian's, Malta, 2016, pp. 689-693, doi: 10.1109/CoDIT.2016.7593646.

[62] Gunawardane, P.D.S.H.; and Medagedara, N.T. "Comparison of hand gesture inputs of leap motion controller & data glove into a soft finger," *2017 IEEE International Symposium on*

Robotics and Intelligent Sensors (IRIS), Ottawa, ON, Canada, 2017, pp. 62-68, doi: 10.1109/IRIS.2017.8250099.

[63] Marin, G.; Dominio, F.; and Zanuttigh, P. "Hand gesture recognition with leap motion and kinect devices," *2014 IEEE International Conference on Image Processing (ICIP)*, Paris, France, 2014, pp. 1565-1569, doi: 10.1109/ICIP.2014.7025313.

[64] Ali, N.M.; and Ali, M.S.M. "Evaluation of Students' Acceptance of the Leap Motion Hand Gesture Application in Teaching Biochemistry," *2019 2nd International Conference on new Trends in Computing Sciences (ICTCS)*, Amman, Jordan, 2019, pp. 1-6, doi: 10.1109/ICTCS.2019.8923107.

[65] Soares, F.; Esteves, J.S.; Carvalho, V.; Lopes, G.; Barbosa, F.; and Ribeiro, P. "Development of a serious game for Portuguese Sign Language," *2015 7th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, Brno, Czech Republic, 2015, pp. 226-230, doi: 10.1109/ICUMT.2015.7382432.

[66] Othman, A.; and El Ghoul, O. "Intra-linguistic and extra-linguistic annotation tool for the “Jumla Dataset” in Qatari sign language," *2021 8th International Conference on ICT & Accessibility (ICTA)*, Tunis, Tunisia, 2021, pp. 01-06, doi: 10.1109/ICTA54582.2021.9809778.

[67] Anwar, A.; Basuki, A.; Sigit, R.; Rahagiyanto, A.; and Zikky, M. "Feature Extraction for Indonesian Sign Language (SIBI) Using Leap Motion Controller," *2017 21st International Computer Science and Engineering Conference (ICSEC)*, Bangkok, Thailand, 2017, pp. 1-5, doi: 10.1109/ICSEC.2017.8443926.

[68] Hein, Z.; Htoo, T.P.; Aye, B.; Htet, S.M.; and Ye, K.Z. "Leap Motion based Myanmar Sign Language Recognition using Machine Learning," *2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus)*, St. Petersburg, Moscow, Russia, 2021, pp. 2304-2310, doi: 10.1109/ElConRus51938.2021.9396496.

[69] Kim, M.; Choi, S.-K.; Yoon, A.; and Kwon, H.-C. "Practical Target Word Selection Method in Korean to Korean Sign Language Machine Translation," *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, Liverpool, UK, 2015, pp. 931-934, doi: 10.1109/CIT/IUCC/DASC/PICOM.2015.140.

[70] Wu, C.-H.; Chiu, Y.-C.; and Guo, C.-S. "Text generation from Taiwanese sign language using a PST-based language model for augmentative communication," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 12, no. 4, pp. 441-454, Dec. 2004, doi: 10.1109/TNSRE.2003.819930.

[71] Madrid, G.K.R.; Villanueva, R.G.R.; and Caya, M.V.C. "Recognition of Dynamic Filipino Sign Language using MediaPipe and Long Short-Term Memory," *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, 2022, pp. 1-6, doi: 10.1109/ICCCNT54827.2022.9984599.

[72] Shokoori, A.F.; Shinwari, M.; Popal, J.A.; and Meena, J. "Sign Language Recognition and Translation into Pashto Language Alphabets," *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, 2022, pp. 1401-1405, doi: 10.1109/ICCMC53470.2022.9753959.

[73] Engineer Ambitiously—NI. Available online: https://www.ni.com/en-gb.html (accessed on 23 October 2022).

[74] Martinek, P.; Krammer, O. Analysing machine learning techniques for predicting the hole-filling in pin-in-paste technology. *Comput. Ind. Eng.* 2019, *136*, 187–194. https://doi.org/10.1016/j.cie.2019.07.033.

[75] Sapkota, S.; Mehdy, A.N.; Reese, S.; Mehrpouyan, H. Falcon: Framework for anomaly detection in industrial control systems. *Electronics* 2020, *9*, 1192. https://doi.org/10.3390/electronics9081192.

[76] Welcome to Python. Available online: https://www.python.org/ (accessed on 19 June 2023).

[77] Dorochowicz, A.; Kurowski, A.; Kostek, B. Employing Subjective Tests and Deep Learning for Discovering the Relationship between Personality Types and Preferred Music Genres. *Electronics* 2020, *9*, 2016.

[78] ASL (American Sign Language). Kaggle Dataset. Available online (accessed on 23 October 2022): https://www.kaggle.com/datasets/alexalex1211/aslamerican-sign-language

[79] Spanish Sign Language Alphabet (Static). Kaggle Dataset. Available online (accessed on 3 January 2022): https://www.kaggle.com/datasets/kirlelea/spanish-sign-language-alphabet-static

[80] Urdu Speech Dataset. Kaggle Dataset. Available online (accessed on 3 January 2022): https://www.kaggle.com/datasets/hazrat/urdu-speech-dataset

Appendix A

Curriculum Vitae

Experience

- Assistant Professor, SSUET — Oct 2008 - date
- Lecturer with additional duty of Network, SSUET — Sep 2006 – Oct 2008
- Demonstrator with additional duty of Network, SSUET — Aug 2004 – Sep 2006
- Research Assistant with additional duty of Network, SSUET — July 2001 – Aug 2004

Education

- Ph.D. Telecommunication Engineering (on progress, expected 2023), Universidad de Málaga.
- MS Computer Engineering with Specialization in Computer Network, SSUET.
- BS Electronic Engineering, SSUET.

Journal Publications

- Saleem, M.I.; Siddiqui, A.; Noor, S.; Luque-Nieto, M.-A.; Nava-Baro, E. A Machine Learning Based Full Duplex System Supporting Multiple Sign Languages for the Deaf and Mute. *Appl. Sci.* **2023**, *13*, 3114. https://doi.org/10.3390/app13053114.
- Saleem, M.I.; Siddiqui, A.; Noor, S.; Luque-Nieto, M.-A.; Otero, P. A Novel MachineLearning Based Two-Way Communication System for Deaf and Mute. *Appl. Sci.* **2023**, *13*, 453. https://doi.org/10.3390/app13010453.
- Saleem, M.I.; Khan, A.M.; Noor, S.; and Aamir, M. "Framework for Smart E-health Monitoring System", *Indian journal of Science and Technology*, Vol. 10, issue 29, August **2017**.

Conference Publications

- Noor, S.; Minhas, H.N.; Saleem, M.I.; Uddin, V.; and Ismat, N. "Inside-out Vision for Procedure Recognition in Dental Environment," Global Conference on Wireless & Optical Technologies 2020 (GCWOT'20), October 2020.
- Saleem, M.I.; Otero, P.; Noor, S.; and Aftab, R. "Full duplex smart for deaf and dumb and normal people," Global Conference on Wireless & Optical Technologies 2020 (GCWOT'20), 6-8 October 2020.
- Saleem, M.I.; Khan, A.M.; Noor, S.; and Aamir, M. "Framework for Smart E-health Monitoring System," 3rd International Conference on Green Computing and Engineering Technologies (ICGCET-2017), 8 - 10 August 2017.