

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
CIUDAD - CHILE



“OPTIMIZACIÓN DE PROTOTIPO DE RECONOCIMIENTO
DE GESTOS FACIALES, BASADO EN VISIÓN
COMPUTACIONAL, PARA LA ASISTENCIA DE PACIENTES
CON DISCAPACIDAD MOTORA SEVERA”

NICOLÁS MARCELO TAPIA MUÑOZ

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN INFORMÁTICA

Profesor Guía: Oscar Reyes H.
Profesor Correferente: Claudio Tapia M.

Diciembre - 2024



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): Memoria o trabajo de título; Tesis de Postgrado;

Título del trabajo: OPTIMIZACIÓN DE PROTOTIPO DE RECONOCIMIENTO

DE GESTOS FACIALES, BASADO EN VISIÓN

COMPUTACIONAL, PARA LA ASISTENCIA DE PACIENTES

CON DISCAPACIDAD MOTORA SEVERA

Nombre del candidato(a): Nicolás Tapia Muñoz

Carrera / Grado: Ingeniería Civil Informática

Campus: Casa Central Valparaíso ; Departamento: Informática

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Oscar Reyes Hevia, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL

El trabajo **NO** contiene información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (embargo) por:

6 meses; 12 meses; 2 años; 3 años; 5 años; 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):
N/A

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 01/06/2025

; Firma:

Estudiante o Candidato(a):

Fecha:

; Firma:

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

DEDICATORIA

Me gustaría dedicar el presente documento a todos aquellos trabajadores y sus familias a quienes la investigación realizada tiene la posibilidad de ayudar, con la esperanza de que contribuya a mejorar sus vidas.

AGRADECIMIENTOS

Agradezco a mi padre primero como profesor y académico por brindarme un lugar para poder realizar mi proyecto de memoria dentro del alcance de uno de sus proyectos y la guía que me entregó a lo largo del desarrollo. También me gustaría agradecerle a él y a mi madre por todo el apoyo incondicional que me han brindado a lo largo de mi carrera universitaria para que pueda terminar.

Agradezco también a Fernando Bernuy quien me acompañó a lo largo del desarrollo, enseñándome nuevas tecnologías con las cuales no estaba familiarizado, ya que no solo fueron puestas en práctica en el presente proyecto, sino que también me han permitido tener un buen desempeño a nivel laboral.

Me gustaría agradecer a mis amigos y a mi pareja por todo el apoyo que han brindado en un sinnúmero de aspectos, académicos y personales.

Y finalmente me gustaría extenderle un agradecimiento especial a los profesores, académicos, funcionarios y estudiantes de la facultad de medicina de la Universidad de Chile, sin quienes no habría una base de datos inicial para el proyecto.

RESUMEN

Resumen— El presente estudio explora la creación de un nuevo modelo de detección de gestos faciales basándose en la detección de expresiones faciales de mediapipe, junto con una plataforma para la creación de una base de datos destinada a su entrenamiento. Además, se abordan todas las configuraciones necesarias para poder servir la plataforma, API y base de datos, y los experimentos realizados con distintas combinaciones de componentes del modelo. El objetivo principal del proyecto es la implementación del modelo en la habitación de un paciente con discapacidad motora severa, permitiéndole realizar acciones simples como el cambio de canal en un televisor, y así aumentar su grado de autonomía. Se abordó el problema con el objetivo de generar una base de datos robusta, enfoque que permitió implementar distintas combinaciones de prácticas de preprocesamiento de datos. El resultado muestra un gran potencial y queda abierto a mejoras con las herramientas desarrolladas en el presente trabajo que apunten a mejorar la precisión del detector de gestos faciales en el futuro a medida que la base de datos crezca.

Palabras Clave— Discapacidad, Autonomía, Nube, Clasificación, Base de Datos.

ABSTRACT

Abstract— This study explores the creation of a new facial gesture detection model based on Mediapipe facial expression detection, along with a platform for creating a database intended for its training. Additionally, all necessary configurations for serving the platform, API, and database are addressed, as well as the experiments conducted with various combinations of model components. The main objective of the project is the implementation of the model in the room of a patient with severe motor disabilities, enabling them to perform simple actions such as changing the television channel, thereby increasing their degree of autonomy. The problem was approached with the aim of generating a robust database, an approach that allowed the implementation of different combinations of data preprocessing techniques. The results show significant potential and remain open to improvements using the tools developed in this work to enhance the precision of the facial gesture detector in the future as the database grows.

Keywords— Disability, Autonomy, Classification, Database, Cloud.

GLOSARIO

HSP: Hosting service Provider. Corresponde a una compañía de servicios TI que provee infraestructura TI remota o en la nube

AWS: Amazon Web Services. Servicio de Amazon que provee soluciones de computación en la nube

SVD: Singular Value Decomposition o Descomposición de Valores Singulares por sus siglas en inglés.

PCA: Principal Component Analysis o Análisis de Componentes principales por sus siglas en inglés.

Frame: Referente a un fotograma de un video.

FPS: Frames per second. Unidad de medida de fotogramas por segundo en video.

JSON: JavaScript Object Notation.

MSE: Mean Squared error o Error Cuadrático Medio.

RF: Random forest

GB: Gradient Boosting

XGB: XGBoost

ÍNDICE DE CONTENIDOS

RESUMEN	IV
ABSTRACT	IV
GLOSARIO	VI
ÍNDICE DE FIGURAS	IX
ÍNDICE DE TABLAS	IX
INTRODUCCIÓN	1
CAPÍTULO 1: DEFINICIÓN DEL PROBLEMA	2
1.1 Contexto	2
1.2 Fase Anterior del Proyecto	5
1.3 Objetivos	5
1.3.1 General	5
1.3.2 Específicos	5
CAPÍTULO 2: MARCO CONCEPTUAL	6
2.1 Discapacidad	6
2.2 Hospitalización	7
2.3 Soluciones Actuales	8
2.4 Soluciones no Exploradas	10
CAPÍTULO 3: PROPUESTA DE SOLUCIÓN	11
3.1 Consideraciones previas	12
3.2 Captura de datos	12
3.2.1 Estructura del Front-end	13
3.2.2 Flujo de uso Front-end	16
3.2.3 Puntos técnicos Front-end	16
3.3 Back-end	17
3.3.1 Rutas de Back-End	17
3.4 Paso a Producción Back-End	18
3.4.1 Etapa 0: Selección de tecnologías	18
3.4.2 Primera Etapa: Levantamiento de Máquina	20
3.4.3 Conexión a Internet	21
3.5 Nuevo Modelo	22
3.6 Recolección de datos	22
3.7 Análisis experimental de modelos de aprendizaje automático	23
3.7.1 Granularidad	23
3.7.2 Reducción de dimensionalidad	23
3.7.3 Regresor	24

3.8 Clasificador	24
CAPÍTULO 4: VALIDACIÓN DE LA SOLUCIÓN	26
4.1 Cantidad de frames	27
4.2 Cantidad de componentes	33
4.3 Clasificador	44
4.3.1 Análisis	47
CAPÍTULO 5: CONCLUSIONES	52
ANEXOS	54
REFERENCIAS BIBLIOGRÁFICAS	63

ÍNDICE DE FIGURAS

1	: Paciente con hospitalización clínica en la Mutual de Seguridad CChC con alternativa a cuff.	8
2	: Blowbell.	9
3	: Cintillo EMG.	10
4	: Propuesta de solución.	11
5	Estructura base de aplicación React.	14
6	: Menú de gestos	15
7	: Inbound rules de Máquina EC2	21
8	Error cuadrático medio para los datos con $F = 5$.	28
9	Error cuadrático medio para los datos con $F = 3$.	29
10	Error cuadrático medio para los datos con $F = 2$.	30
11	Error cuadrático medio para datos con $F = 1$.	31
12	Comparación de MSE con F para técnicas de reducción de dimensionalidad	32
13	Error cuadrático medio para datos con $C = 2$.	33
14	Varianza explicada de PCA con $C = 2$	34
15	Influencia de los componentes SVD con $C = 2$	35
16	Error cuadrático medio para datos con $C = 2$.	36
17	Varianza explicada de PCA con $C = 3$	37
18	Influencia de los componentes SVD con $C = 3$	38
19	Varianza explicada de PCA con $C = 5$	39
20	Influencia de los componentes SVD con $C = 5$	40
21	Error cuadrático medio para datos con $C = 2$.	41
22	Varianza explicada de PCA con $C = 7$	42
23	Influencia de los componentes SVD con $C = 7$	42
24	Grafico de Cantidad de Componentes vs tiempo que demoró entrenar el modelo	43
25	Grafico de Cantidad de Componentes vs tiempo que demoró entrenar el modelo	44
26	Matriz de confusión para $C = 2$	47
27	Matriz de confusión para $C = 3$	48
28	Matriz de confusión para $C = 4$	49
29	Accuracy vs cantidad de componentes principales	50
30	Matriz de confusión para $C = 9$	51
31	: Vista de Formulario.	54
32	: Vista de menú.	62
33	: Mensaje de salida.	62

ÍNDICE DE TABLAS

1	Especificaciones de la Máquina	19
2	Estructura general del dataframe de los datos.	26

3	Resultados de Modelos para $F = 5$	28
4	Resultados de Modelos para $F = 3$	29
5	Resultados de Modelos para $F = 2$	30
6	Error cuadrático medio para los datos con $F = 1$.	31
7	Error cuadrático medio para los datos con $C = 2$.	33
8	Error cuadrático medio para los datos con $C = 3$.	35
9	Error cuadrático medio para los datos con $C = 7$.	41
10	Asignaciones de encoding para expresiones	45
11	Distribución de etiquetas en los datos de entrenamiento	46
12	Resultados de precisión y recall por etiqueta para $C = 2$	48
13	Resultados de precisión y recall por etiqueta para $C = 3$	49
14	Resultados de precisión y recall por etiqueta para $C = 4$	50
15	Resultados de precisión y recall por etiqueta	51
16	Lista de Expresiones	63

INTRODUCCIÓN

La integración de tecnologías en los procesos y prácticas de distintas instituciones es conocida como transformación digital, la cual puede ser implementada en diversos ambientes laborales o académicos. Particularmente en el área médica, la transformación digital puede aportar valor en muchos aspectos, como lo puede ser Mantener sistemas de bajo costo, ecosistemas digitales y en general poblaciones más saludables [World Health Organization, 2021]. El objetivo de este documento es la creación de un modelo de clasificación de gestos faciales para la segunda etapa del proyecto *Desarrollo de un sistema de reconocimiento de gestos faciales basado en visión computacional para la asistencia de pacientes con discapacidad motora severa crónica de Mutual de Seguridad CChC.*, el cual tiene como fin brindar un mayor grado de autonomía a los pacientes con discapacidad motora severa, quienes a día de hoy dependen de sus cuidadores o enfermeras dependiendo de su situación de hospitalización. Tal grado de dependencia afecta no solo al paciente sino también a quienes los rodean. Sin embargo, con la implementación del sistema propuesto en el proyecto con la capacidad de detectar los gestos faciales realizados por el paciente, este podrá controlar aspectos de su habitación con facilidad, por ejemplo el encendido y apagado de las luces, televisión o radio.

Para generar este nuevo clasificador de gestos faciales se tomó en cuenta la posibilidad de aumentar los datos de entrenamiento a lo largo del tiempo, por lo que fue desarrollado un portal web que permita guardar los datos de los usuarios y pacientes tras la firma digital de un consentimiento informado. Luego se empleó la biblioteca mediapipe proporcionada por Google para poder extraer las expresiones realizadas por las personas, para posteriormente entrenar un regresor, que alimenta a un clasificador con los componentes principales obtenidos de una reducción de dimensionalidad de los datos.

A continuación, se aborda una comparación de como distintos factores pueden afectar el resultado del clasificador final, como lo son la cantidad de frames tomados del video, la cantidad de componentes principales considerados, la técnica de reducción de dimensionalidad empleada y el regresor que presenta el mejor rendimiento.

Para finalizar, se presentan las conclusiones y el trabajo a futuro planteado para siguientes iteraciones del mismo proyecto de asistencia a pacientes.

CAPÍTULO 1

DEFINICIÓN DEL PROBLEMA

1.1. Contexto

La presente memoria será realizada como parte del proyecto Implementación clínica y optimización de prototipo de reconocimiento de gestos faciales, basado en visión computacional, para la asistencia de pacientes con discapacidad motora severa de Mutual de Seguridad, Claudio Tapia Malebrán, PHD. y Carlos Cruz Montecinos, Mgtr., adjudicado a la Universidad de Chile por parte de la Mutual de Seguridad CChC.

Según la definición de la OMS de discapacidad, esta es un fenómeno complejo que refleja una relación estrecha y al límite entre las características del ser humano y las características del entorno en donde vive. Y según la Encuesta Nacional de Discapacidad ENDISC se entiende como una construcción simbólica, un término genérico y relacional que incluye condiciones de salud y déficits, limitaciones en la actividad, y restricciones en la participación. Este concepto indica los aspectos negativos de la interacción, entre un individuo y sus factores contextuales, considerando los ambientales y personales. Además, este mismo define a las personas con discapacidad como aquellas personas que, en relación con sus condiciones de salud física, psíquica, intelectual, sensorial u otras, al interactuar con diversas barreras contextuales, actitudinales y ambientales, presentan restricciones en su participación plena y activa en la sociedad.

La OMS, además, identifica cuatro distintos factores que contribuyen a la desigualdad con relación a la salud:

- Factores estructurales: Las personas con discapacidad experimentan el capacitismo, la estigmatización y la discriminación en todas las facetas de su vida, lo que afecta a su salud física y mental.
- Determinantes sociales: La pobreza, la exclusión de la educación y el empleo, y las malas condiciones de vida aumentan el riesgo de que las personas con discapacidad padezcan mala salud y no tengan cubiertas sus necesidades en este terreno.
- Factores de riesgos: Las personas con discapacidad tienen más probabilidades de presentar factores de riesgo de enfermedades no transmisibles, como el tabaquismo, la mala alimentación, el consumo de alcohol y la falta de actividad física.
- Sistemas de salud: Las personas con discapacidad se enfrentan a barreras en todos los aspectos del sistema de salud. Por ejemplo, falta de conocimientos, actitudes negativas y prácticas discriminatorias entre el personal de atención de salud; instalaciones e información inaccesibles; y falta de información o de recopilación y análisis de datos sobre discapacidad.

El proyecto aborda la población con discapacidad motora severa crónica, quienes ven afectada su autonomía, independencia, participación y muchas veces presentan una gran carga emocional asociada a su condición de salud y dependencia total de terceras personas. Busca potenciar la autonomía de las y los trabajadores a través del desarrollo de un prototipo basado en el uso de visión computacional e inteligencia artificial que reconoce gestos faciales con el fin de controlar diferentes dispositivos disponibles en la habitación del paciente.

De acuerdo con el II Estudio Nacional de Discapacidad (ENDISC) de 2015, se estima que el 20 % de la población adulta chilena posee algún grado de discapacidad, de esta el 8,3 % posee discapacidad severa, y de esta un 75,7 % de la población inactiva en el ámbito laboral. Estos números se mantuvieron similares en el III ENDISC de 2022, en donde la población con algún grado de discapacidad disminuyó a un 17,6 %, lo que corresponde a 2.703.893 personas estimadas frente a las 2.606.914 del II ENDISC por lo que el número total aumentó.

Una de las principales condiciones de salud que genera discapacidad severa y crónica es la lesión de médula espinal la cual hace referencia a los daños sufridos en la médula espinal a consecuencia de un traumatismo (por ejemplo, un accidente de coche) o de una enfermedad o degeneración (por ejemplo, el cáncer). No existen estimaciones fiables de su prevalencia mundial, pero se calcula que su incidencia mundial anual oscila entre 40 y 80 casos por millón de habitantes. Hasta un 90 % de esos casos se deben a causas traumáticas, aunque la proporción de lesiones medulares de origen no traumático parece ir en aumento (WHO, 2013). De acuerdo con la Organización Mundial de la Salud, entre 250.000 y 500.000 personas sufren cada año lesiones medulares en todo el mundo, la mayoría se deben a causas traumáticas prevenibles como accidentes de tránsito, caídas o actos de violencia (WHO, 2011).

Las personas con lesiones medulares son entre dos y cinco veces más propensas a morir prematuramente que las que no los padecen, con el riesgo de mortalidad alcanzando su nivel máximo en el primer año después de sufrir la lesión, y manteniéndose elevado en comparación con la población general en los años siguientes. El riesgo de mortalidad aumenta en función de la localización y la gravedad de la lesión y varía sobremanera según se dispense o no oportunamente una atención médica de calidad. El método utilizado para trasladar al hospital a la persona que acaba de sufrir una lesión de este tipo y el tiempo que tarda en ingresar son también factores importantes.

Particularmente el traumatismo raquimedular (TRM) resulta de la lesión de estructuras óseas, ligamentosas y/o neurológicas de la columna vertebral, afectando la conducción de las señales motoras y sensitivas en el o los sitios de la lesión además del sistema nervioso autónomo, generando deterioro motor y diferentes niveles de discapacidad.

Distintos reportes han descrito a nivel mundial que la incidencia de un TRM alcanza 10,5 casos por 100.000 personas, aumentando en países con menores ingresos, llegando a 13,7 casos por 100.000, mientras que en países con ingresos altos esta relación disminuye a 8,7 casos por 100.000 personas. En Chile, de acuerdo al Global Burden Disease del Institute for Health Metrics and Evaluation de la Universidad de Washington, en el año 2017 se reportó una prevalencia de 51.616 con una incidencia de 1.618 casos.

Esta condición afecta más a hombres que mujeres, se sugiere que esto puede estar determinado por peligros ocupacionales únicos o comportamientos más riesgosos que los hacen vulnerables a traumas. Si se considera que el promedio de edad, a nivel mundial, de las personas que presentan esta condición es de 40 años, resulta evidente el impacto que esto tiene a nivel laboral. En estricto rigor, la principal causa de TRM corresponde a los accidentes de tránsito, llegando a 41,6 % de los casos reportados, seguido por caídas, las que ocurren principalmente en el trabajo (WHO, 2011).

Esta condición afecta el bienestar físico, psicológico y social del paciente. Por lo tanto, mejorar el acceso y calidad de la atención médica y el acceso a rehabilitación oportuna, mejora la supervivencia y reduce la morbilidad, discapacidad y dependencia. Un estudio realizado en pacientes con discapacidad producto de lesión medular en Chile describe la vivencia de tener una discapacidad por daño a la medula espinal, lo difícil que es asumir esta condición y disponerse a la rehabilitación, describiendo diferentes etapas en el proceso de rehabilitación; desde una fase marcada por la depresión, hasta alcanzar en los mejores casos la independencia absoluta. Desde este punto de vista, resulta pertinente enfatizar la discapacidad y dependencia a la que se enfrenta una persona con lesión medular puesto que una persona con lesión medular y con secuelas motoras severas podría ver impedida su participación y autonomía, generando un aumento del nivel y percepción de dependencia, aumentando con ello su discapacidad.

Las lesiones medulares o condiciones de salud que impliquen secuela de dependencia severa aumentan también significativamente los costos hospitalarios, generando una carga financiera sustancial en los sistemas de atención médica asociado a los altos costos de atención de la salud en esta población. Además, de manera indirecta, esta carga financiera del sistema se agrava producto de la pérdida de productividad. El manejo del TRM requiere de importantes recursos sanitarios y puede representar una carga financiera considerable para los pacientes, sus familias y la comunidad. Estos costos se deben, en gran medida, a la necesidad de cuidados agudos a corto plazo en la Unidad de Paciente Crítico (UPC) y las complicaciones secundarias asociadas que ocurren a corto, mediano y largo plazo (MINSAL, 2019). La pérdida de autonomía, con la consiguiente discapacidad asociada lo exponen a un aumento del riesgo de dichas complicaciones, entre las que se encuentran: las fracturas asociadas a osteoporosis por desuso, compromiso de funciones vitales, atrofia muscular y la presencia de úlceras por presión (UPP). Por ejemplo, se estima que en EE. UU. los costos asociados debido al manejo de UPP se aproximan a los 11 billones de dólares por año, mientras que en Europa se estima que corresponden el 1,4 % del gasto en salud.

Finalmente, el problema de autonomía de un paciente con discapacidad motora severa crónica tiene diversas aristas que son necesarias enfrentar hoy en día como lo son el costo monetario y humano de brindarle asistencia a los pacientes tanto para la familia como para la Mutual de Seguridad, el efecto psicológico que tiene su condición sobre él mismo y sobre su red de apoyo, la falta de autonomía que le permita desenvolverse en su entorno, entre otros.

1.2. Fase Anterior del Proyecto

La fase anterior tuvo como objetivo general “Desarrollar un sistema de reconocimiento de gestos faciales basado en visión computacional para la asistencia de pacientes con discapacidad motora severa crónica de Mutual de Seguridad CChC”. Las etapas implementadas fueron cuatro:

1. Desarrollo de software de detección de puntos de interés de rostro mediante el procesamiento de imágenes de video.
2. Entrenamiento de algoritmos de detección de patrones gestuales basado en características individuales.
3. Desarrollo de comunicación para control de dispositivo en tiempo real.
4. Validación clínica de sistema de reconocimiento de gestos faciales basado en visión computacional para la asistencia en tiempo real de pacientes con discapacidad motora severa crónica de Mutual de Seguridad CChC.

1.3. Objetivos

Se presentan los siguientes objetivos

1.3.1. General

Optimizar el prototipo desarrollado en la etapa de iniciación para validar el sistema de reconocimiento de gestos faciales, basado en visión computacional, usado en la asistencia y acompañamiento de pacientes con discapacidad motora severa hospitalizados en Mutual de Seguridad o con hospitalización domiciliaria.

1.3.2. Específicos

- I Crear una aplicación que se encargue de grabar y almacenar de manera automática los videos y datos de los usuarios y pacientes que colaborarán con la base de datos.
- II Comparar distintas combinaciones de metaheurísticas para obtener el mejor resultado del clasificador.
- III Validación del nuevo modelo.

CAPÍTULO 2

MARCO CONCEPTUAL

2.1. Discapacidad

La discapacidad, como fue visto anteriormente, depende de la relación de una persona con su entorno, y cuanta ayuda necesita para relacionarse con este. La legislación chilena define distintos grados de discapacidad, conforme a las siguientes definiciones:

- a) Persona sin discapacidad: Es aquella que no presenta limitaciones para realizar actividades propias de su edad y/o no presenta restricciones a la participación, o bien, estas limitaciones y/o restricciones se presentan en un rango entre 0 % a 4 % a causa de su condición de salud.
- b) Persona con discapacidad leve: Es aquella que presenta entre 5 % y 24 % de restricciones en la participación o limitaciones en las actividades propias de su edad a causa de sus condiciones de salud.
- c) Persona con discapacidad moderada: Es aquella que presenta entre 25 % y 49 % de restricciones en la participación o limitaciones en las actividades propias de su edad a causa de sus condiciones de salud.
- d) Persona con discapacidad severa: Es aquella que presenta entre 50 % y 94 % de restricciones en la participación o limitaciones en las actividades propias de su edad a causa de sus condiciones de salud.
- e) Persona con discapacidad profunda: Es aquella que presenta entre 95 % y 100 % de restricciones en la participación o limitaciones en las actividades propias de su edad a causa de sus condiciones de salud.

Para casos de esta memoria, las personas que son de interés son aquellas con discapacidad motora severa crónica, lo que significa que son aquellas personas que tienen una discapacidad severa, relacionada a su motricidad, que no se pueden recuperar de su condición. Al presentar una discapacidad motora, pueden comunicarse de manera verbal o no verbal - esta segunda siendo la única opción para pacientes con traqueotomía¹ ya que no presentan compromiso cerebral. Además, se reduce a los pacientes de la Mutual de Seguridad CChC.

¹Apertura en frente del cuello que se hace durante un procedimiento de emergencia o una cirugía planeada

2.2. Hospitalización

Los pacientes con tetraplejia (con movilidad solamente desde el cuello hacia abajo) pueden encontrarse dentro de dos categorías dependiendo de su estado hospitalario:

- a) Hospitalización clínica: Estos pacientes están dentro del hospital en el área de neurología, son de larga estadía ya que son pacientes crónicos, pero están bajo un protocolo normal de hospitalización donde pueden recibir visitas en los horarios especificados, tienen atención de enfermeras de manera rutinaria, y deben de compartir una habitación con otro paciente. Cada habitación tiene 2 camas por lo que un paciente puede tener que compartir su habitación, pero cada uno tiene su propio sistema de monitoreo que consta de monitor de presión arterial, monitor de oxigenación, de presión, etc. y su propia red de oxígeno (como puede ser visto en la Figura 1).
- b) Hospitalización domiciliaria: La hospitalización domiciliaria va a variar en cada caso dependiendo de muchos factores, entre los que se encuentran la situación socioeconómica, ubicación geográfica, red de apoyo, etc. Como será visto un poco más adelante los pacientes requieren de ciertos cuidados específicos los cuales pueden ser llevados a cabo por profesionales de la salud que debe enviar la mutual de seguridad al domicilio del paciente, y/o por miembros del núcleo familiar, o encargados que puede contratar la familia. A pesar de eso existen casos poco comunes como pacientes que residen en poblaciones de riesgo a donde los profesionales no pueden ir por seguridad propia, o pacientes en donde la familia toma el cargo de toda la asistencia que necesite el paciente, para estos últimos el estado debe entregarles un bono, pero no son la norma estos casos.

Estos pacientes son completamente dependientes motores, lo que significa que necesitan ayuda en todas sus acciones y/o actividades. Requieren de cuidados y cuidadores que debe proveer la Mutual de Seguridad CChC por ley, entre los que se encuentran:

- Cuidador encargado de cuidados varios, como de mover al paciente para evitar exco-riaciones² y lesiones por presión. Esto se hace una vez cada dos horas excepto cuando el paciente está durmiendo por protocolo. El cuidador también es el encargado de otorgarle asistencia en lo que necesite, como encender o apagar el televisor, abrir o cerrar las ventanas, abrigarse o desabrigarse, etc. A este cuidador el paciente debe llamarle mediante un cuff que cuelga contra la cara del paciente constantemente en el caso de estar bajo hospitalización clínica, y el cual el paciente debe morder para avisar que requiere asistencia. Además, debe haber una enfermera con técnico nivel superior encargada de sondear a los pacientes, esto es debido a que no se pueden mover. En el caso de hospitalización domiciliaria el cargo de cuidador queda a discreción de la familia, pero en casos particulares pueden pedir a la Mutual (o esta misma decidir que es apropiado enviarlo) la visita de una enfermera, médico u otro especialista.

²Irritación cutánea que se presenta donde la piel roza contra ella misma, las ropas u otro material



Figura 1: : Paciente con hospitalización clínica en la Mutual de Seguridad CChC con alternativa a cuff.

Fuente: Cortesía de Claudio Tapia.

- Kinesiólogo encargado de ayudar al paciente a mover sus articulaciones. Esto es sumamente importante para evitar fracturas al momento de cambiar de posición al paciente debido a osteoporosis³ por desuso. Además, ayuda a prevenir la anquilosis o rigidez de las articulaciones y la atrofia muscular.

Estos dos roles son obligatorios en el cuidado de un paciente con tetraplejía.

Los pacientes además pueden necesitar de asistencia respiratoria, pero esto varía de caso en caso.

2.3. Soluciones Actuales

Actualmente hay 3 maneras con las cuales la mutual de seguridad enfrenta el cuidado de un paciente tetrapléjico. Estas opciones van desde la menos a la más tecnológica.

- i) La opción más básica es tener un cuidador permanentemente al lado de un paciente, lo que es caro y desgastante para el cuidador. Esta opción no se utiliza en la actualidad.
- ii) Un cuidador no permanente a quien el paciente en situación de hospitalización clínica pueda llamar mediante morder el cuff. Esta opción, aunque más efectiva que la anterior, es denigrante para el paciente, ya que debe tener el cuff colgando contra su

³Enfermedad que se caracteriza por una disminución de la densidad de los huesos a causa de la pérdida del tejido óseo normal

cara constantemente. Además, esta opción presenta un foco de infecciones ya que el paciente debe morder el cuff.

- iii) Un blow bell que es una mejora a la opción anterior. Consiste en un tubo en el cual el paciente sopla para poder llamar la atención de la enfermera. Esto se puede apreciar en la figura 2.



Figura 2: : Blowbell.

caregiverlogy. <https://www.caregiverology.com/call-light.html>

- iv) Un cintillo con capacidades de electromiografía (EMG) que lee constantemente los impulsos nerviosos de las cejas del paciente permitiéndole así controlar ciertos aspectos, como el volumen del televisor o llamar a la enfermera. Esta opción presenta un salto gigantesco en la autonomía del paciente ya que le permite realizar acciones simples sin necesidad de asistencia. Esta opción era la mejor y más moderna, pero debido

a la pandemia la Mutual de Seguridad decidió disminuir el contacto entre pacientes por obvias razones. Un ejemplo de esto puede ser visto en la figura 3.

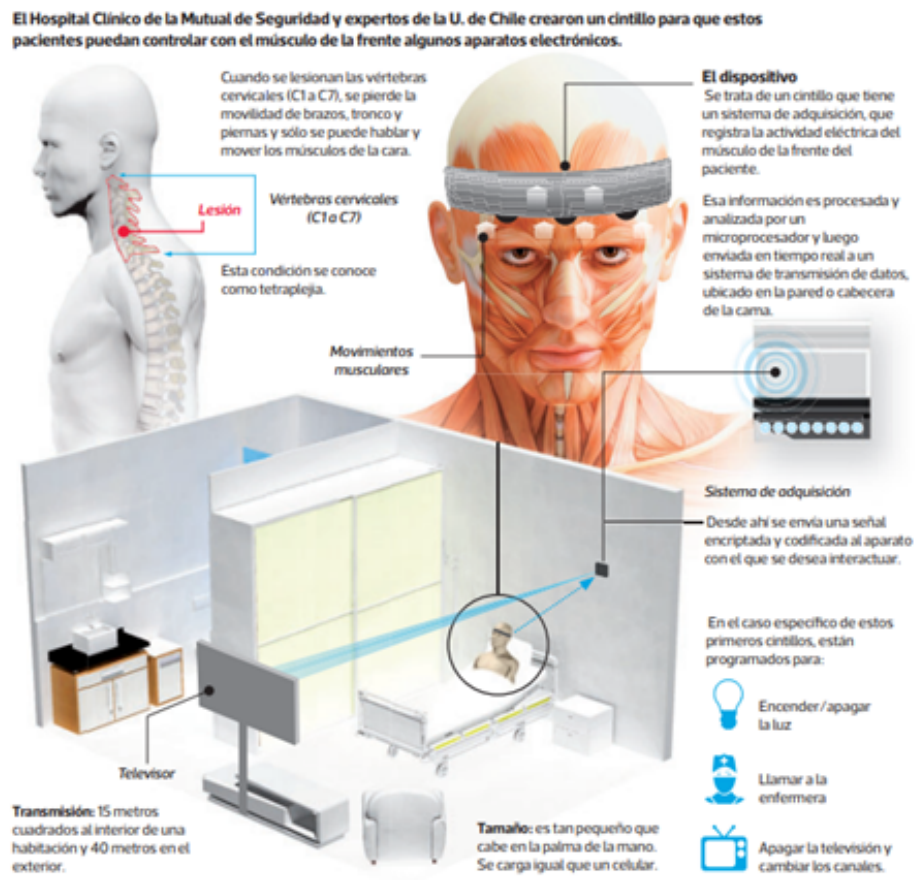


Figura 3: : Cintillo EMG.
Fuente: La Tercera.

2.4. Soluciones no Exploradas

Hoy en día existe un vacío de soluciones que no impliquen contacto físico entre el paciente y la herramienta que utilice para pedir ayuda o que le otorgue algún grado de autonomía, por lo que explorar este ámbito utilizando nuevas tecnologías se visualiza como novedoso e imperativo.

CAPÍTULO 3 PROPUESTA DE SOLUCIÓN

La solución que se propone como parte del proyecto es la de una herramienta que utilice visión computacional para identificar movimientos faciales del paciente, mediante los cuales pueda realizar acciones básicas como encender el televisor o apagar las luces, sin necesidad de tener contacto físico con la herramienta.

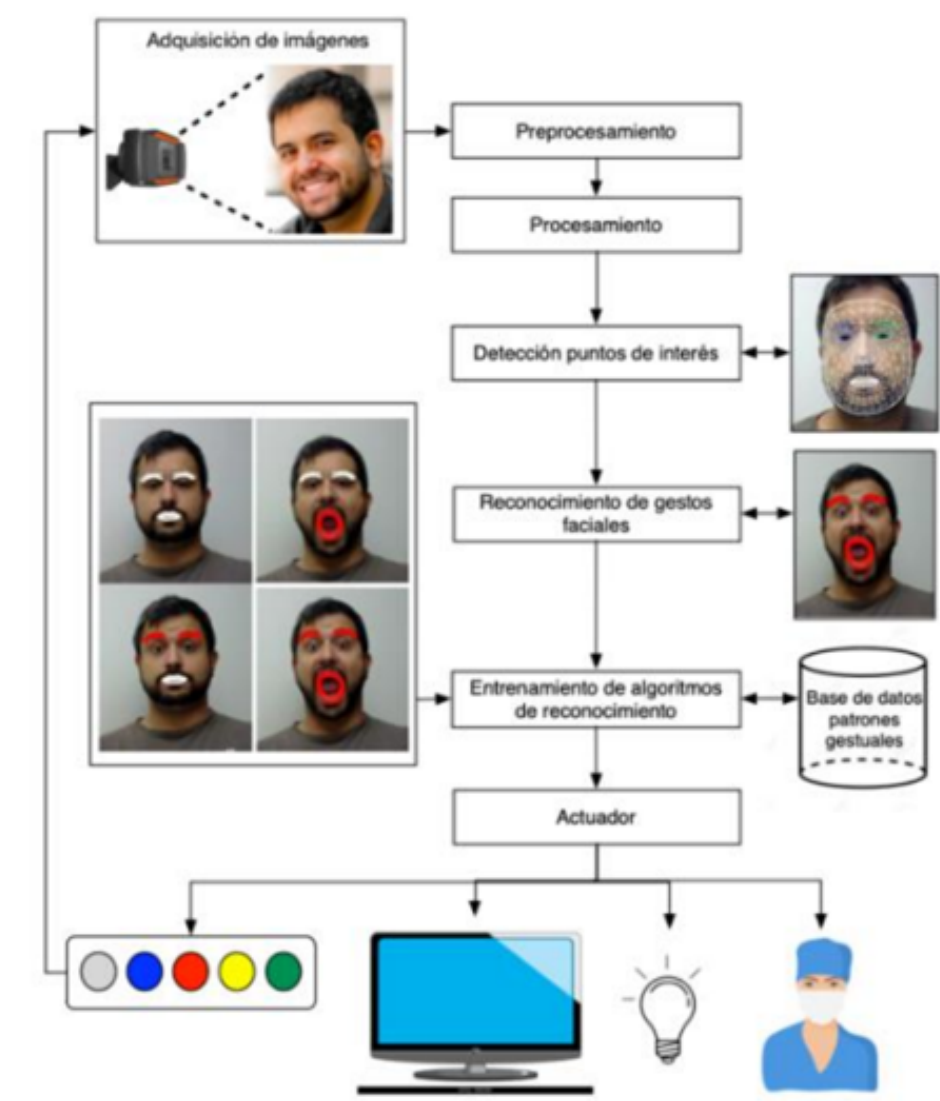


Figura 4: : Propuesta de solución.
Fuente: Cortesía de Claudio Tapia.

El valor agregado de la solución propuesta en contraste con los métodos utilizados hoy en día se puede abordar desde diferentes puntos de vista, entre otros:

- No implica contacto físico con el paciente, situación de vital importancia en entornos hospitalarios, sobre todo considerando los riesgos asociados a la pandemia por Covid-19 y a la condición de salud del paciente.
- Constituye un sistema de asistencia para personas en situación de dependencia motora, aumentando su autonomía.
- Impacta en la salud mental del paciente, pues permite el logro de objetivos sin la asistencia de un tercero.
- Disminuye los costos asociados al cuidado y asistencia de estos pacientes.

Los pasos ideados para llevar a cabo la solución planteada son:

1. Crear una plataforma de captura y almacenamiento de datos.
2. Preprocesar dichos datos utilizando distintos métodos aislados y en combinación.
3. Utilizar los datos preprocesados para alimentar distintos clasificadores.
4. Comparar el rendimiento de cada combinación de datos y clasificadores.

3.1. Consideraciones previas

En vista de que el objetivo de esta memoria es subyacente a un proyecto que será implementado en la Mutual de Seguridad, el equipo de investigación solicitó que el desarrollo se realizara considerando un trabajo a futuro posible que permitiera mejorar en el tiempo el clasificador resultante. Es por esto que en algunos casos como lo serán el desarrollo de la plataforma de toma de datos, se diseñó bajo el pensamiento de poder agregar más datos a la base de datos con la cual será entrenado el clasificador.

3.2. Captura de datos

Como será visto en secciones posteriores, se decidió hacer uso de la biblioteca MediaPipe de Google, la cual entrega buenos resultados al ser usado con personas fenotípicamente estadounidenses (pelo claro, tez de piel clara), pero en versiones anteriores del proyecto, ha tenido problemas detectando personas de color de piel más oscuro, o con rasgos latinos más marcados. Es en base a esta experiencia que se decidió generar una nueva base de datos de personas latinas, ya que esto generará un conjunto de datos representativo del usuario objetivo de la herramienta. Para la captura de datos se optó por utilizar una aplicación web debido a las ventajas que presenta para la facilidad de uso: no se debe instalar una aplicación

desde un mercado o tienda de aplicaciones ⁴ ni se debe entregar el instalador a cada usuario que quiera colaborar con la toma de datos, además esto nos permite acceder a la misma desde un computador, ya sea portátil o de escritorio, como también desde dispositivos móviles como celulares o tabletas. Esto último es importante, ya que ayuda sustancialmente con la captura de datos de los mismos pacientes, quienes no pueden mover más allá de sus cuellos por su estado.

Las funciones básicas de la aplicación web se limitaron a tres:

- La capacidad de almacenar datos del paciente como nombre y edad.
- La capacidad de almacenar una copia del consentimiento informado del paciente con una copia firmada digitalmente.
- la capacidad de grabar videos.

La aplicación web debe estar formada por dos partes: un front-end y un back-end. Estas dos partes deben de poder comunicarse constantemente para poder cumplir con las funciones previamente establecidas. La primera parte, el front-end, sería creada y estructurada utilizando React.js debido a la familiaridad con el framework y, dado que se utilizaría React.js para el frontend, el cual es una biblioteca que viene por defecto en Node.js, se utilizó Node.js para el desarrollo de la segunda parte, el back-end.

3.2.1. Estructura del Front-end

Como fue mencionado anteriormente, el front-end de la aplicación sería desarrollado en React. Al iniciar un proyecto React con create-React-app [[React.js, 2024](#)], Node se encarga de generar la estructura básica para una página web. El objetivo de crear todos los archivos dentro del directorio es entregar todas las herramientas necesarias a los desarrolladores para poder añadir fácilmente nuevas vistas, nuevos componentes, nuevos servicios, etc. La estructura básica de un proyecto React es la mostrada en la figura ⁵

En donde por convención general se genera una carpeta “src/components” dentro de la cual se van agregando todos los componentes creados que pueden ser vistas nuevas, componentes a reutilizar, funciones, etc.

En este caso, dentro de la carpeta components están los archivos de

⁴Un mercado de aplicaciones o tienda de aplicaciones, como los son Google Play o la App Store de IOS, son plataformas de distribución digital que le permiten a los desarrolladores publicar y vender aplicaciones móviles y de escritorio. Ambas plataformas facilitan la descarga e instalación de aplicaciones y ofrecen herramientas para monetización y análisis del rendimiento de las aplicaciones [[Wikipedia, 2024b](#)]

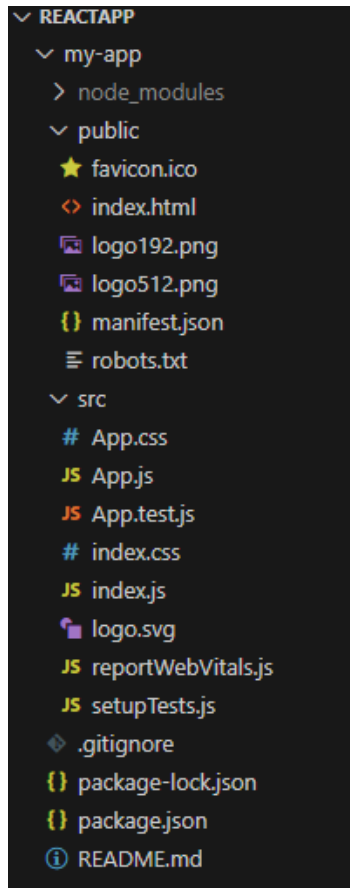


Figura 5: Estructura base de aplicación React.
Fuente: Elaboración propia.

- CameraCapture.js: Este componente contiene la vista y lógica de la captura de video para todos los componentes. Tiene un playback de video para mostrar un ejemplo del movimiento en cuestión, el feedback de la cámara y un segundo playback en donde se puede ver el vídeo más reciente grabado para un gesto.
- SearchView.js: Este componente está fuera del flujo normal de la aplicación y cumple la función de buscar la existencia del consentimiento informado de una persona en específico basándose en su nombre.
- ThankYou.js: Este componente corresponde a una vista de agradecimiento para finalizar el proceso de toma de videos.
- Tos.js: Este componente corresponde al formulario que verá el usuario al entrar a la aplicación, se encarga de tomar los datos de nombre, apellido, edad, sexo, enfermedad de base, peso estimado y altura en centímetros. Además, tiene un visualizador de pdf para mostrar el consentimiento informado y un componente checkbox para confirmar haber leído y estar de acuerdo con el consentimiento informado.

- **UserContext.js**: Este componente crea un contexto del usuario para poder referenciar el nombre del usuario y otros datos de interés durante el proceso, como el timestamp de creación para la carpeta.
- **menu.js**: Este último componente corresponde a un menú que te lleva a las distintas vistas de toma de video, una para cada gesto. Adicionalmente, el menú contiene una pequeña explicación y guía sobre el proceso de toma de videos. El menú puede ser visto en la figura 6.

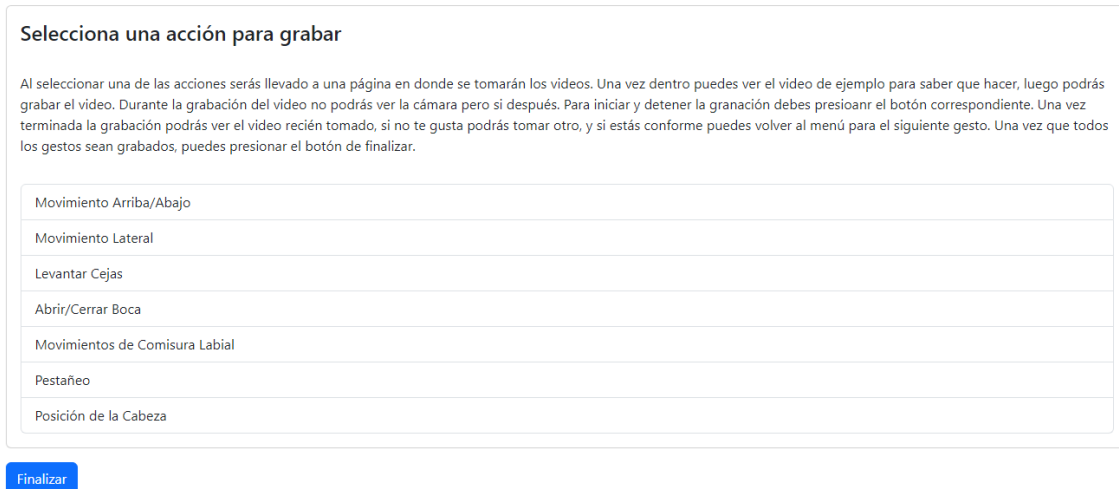


Figura 6: : Menú de gestos
Fuente: Elaboración propia.

Adicionalmente, dentro del mismo directorio /src se creó la carpeta services la cual tendría dentro cualquier servicio que no fuera un componente como tal, sin embargo, solamente se creó un servicio:

- **UploadVideo.js**: este servicio se encarga de crear el payload que se envía a la api con el video, el nombre del archivo y la información del usuario, y se lo envía a la ruta /upload de la api.

También, dentro del mismo directorio /src se creó la carpeta /assets, la cual contiene el PDF del consentimiento informado. Este archivo se copia para cada usuario al momento de ingresar todos sus datos para así tener una firma digital del usuario en forma de un hash creado con el nombre del usuario y la fecha, como fue mencionado anteriormente, el cual

En el directorio raíz del proyecto React también se encuentra la carpeta public, la cual se usa en general para agregar cualquier otro asset esencial para el funcionamiento de la aplicación web, como lo son el favicon (icono que se muestra en la pestaña del navegador [Wikipedia, 2024a]) o el manifest.json (documento JSON que entrega información sobre la aplicación web [MDN, 2024]).

3.2.2. Flujo de uso Front-end

3.2.3. Puntos técnicos Front-end

La página web fue desarrollada utilizando la biblioteca React de Node.js para desarrollo web, además de la biblioteca de diseño Bootstrap 5.

La página cuenta con un total de 5 vistas distintas en cuanto a componentes, pero durante el uso normal de la página, el usuario verá 10 vistas y rutas distintas. Esto se debe a que todas las vistas de toma de videos de gestos reutilizan el mismo componente, el cual cambia por el árbol de rutas y por cómo se llama la vista de esta, esto se hizo para evitar escribir el mismo código para todos los movimientos y tiene el beneficio adicional de permitir la creación de nuevas vistas para gestos objetivos adicionales a los que ya están siendo contemplados. Además, existe una vista creada exclusivamente para revisar que exista efectivamente el consentimiento informado de alguno de los usuarios o pacientes que utilizaron la plataforma.

El pdf de consentimiento informado está como asset en la carpeta “public” del proyecto, al igual que los videos de ejemplo que se muestran antes de grabar un video. Esto permite, junto al uso de bootstrap que la página pueda ser usada tanto en celular como en un computador de escritorio sin ningún problema, con la única diferencia siendo que en un computador el pdf se muestra en un visualizador online, mientras que en dispositivos móviles el documento debe ser descargado.

Debido a que uno de los puntos recurrentes del proyecto es la privacidad de los videos de los usuarios, antes de poder ingresar al menú de selección de gestos se debe confirmar la lectura del consentimiento informado, lo cual crea una copia del mismo en la máquina de la api (más información sobre esto en la siguiente sección).

En cuanto a las vistas de grabación de gestos, se decidió limitar el framerate de la captura a 30 fps y a una resolución de 1280 X 720 pixeles ya que no es necesario tener tanta granularidad en cuanto a cuadros por segundo, y se limitó también la resolución a la cual se graba para poder tener un símil a las cámaras que serán utilizadas con la implementación final del modelo.

La decisión de deshabilitar la vista de la cámara durante la grabación fue tomada por el líder del proyecto debido a que las personas tendemos a exagerar los gestos cuando podemos vernos mientras los estamos haciendo (uno puede realizar el ejercicio de grabarse haciendo el mismo gesto con la cámara delantera y trasera y comparar las diferencias), sin embargo, la implementación final no tendrá feedback para que los pacientes puedan realizar los gestos de la manera más natural para ellos, por lo que esta exageración inconsciente no ocurrirá, debido a esto los datos con los que se entrenó el modelo siguen la misma lógica.

La aplicación React está siendo hosteada por VERCEL, un HSP gratuito para distintas aplicaciones y programas entre los que se encuentran aplicaciones React, APIs REST en Node.js,

aplicaciones dockerizadas, entre muchas más. VERCEL además proporciona nombres de dominio genéricos a sus usuarios (en este caso el nombre que se le fue dado al proyecto React fue memoria-kappa.vercel.app) y, debido a que está vinculado directamente con el repositorio Github, al subir cualquier cambio a la rama main, VERCEL aplica los cambios con un delay de menos de 5 minutos. Todos estos aspectos del servicio de hosting hicieron que el frontend de la página fuera sumamente fácil de subir a internet, mantener, utilizar y conectar con la parte correspondiente al backend de la página.

3.3. Back-end

El backend de la página consta de una api REST Node.js con diversas rutas que se encarga, en términos generales, del guardado de los videos para posterior preprocesamiento y procesamiento, y de guardar la información ingresada en el formulario por los usuarios y pacientes que usen la plataforma.

3.3.1. Rutas de Back-End

Al ser una api un servicio el cual se debe alcanzar mediante rutas y peticiones de disntintos tipos, se crearon las siguientes rutas:

- **/upload**
Método: POST
Función: Esta ruta se encarga de subir los videos a las carpetas correspondiente tomando en cuenta el nombre de usuario y el timestamp de la sesión, además de la vista para el nombre del video.
- **/submit-form**
Método: POST
Función: Esta ruta se encarga de subir los datos del formulario a un csv donde son almacenados todos los datos de los usuarios.
- **/create-user-directory**
Método: POST
Función: Esta ruta se encarga de crear el directorio con el hash creado en base al nombre y el timestamp para el posterior almacenaje de los videos.
- **/search/:name**
Método: GET
Función: Esta ruta se encarga de buscar y retornar el o los consentimientos informados asociados a un nombre.

- /
Método: GET
Función: Esta ruta es solamente de prueba y no es llamada en ninguna parte del código, es solamente depuración para asegurarse de que la api sí esté activa y devuelve el mensaje "¡El servidor está funcionando!".

3.4. Paso a Producción Back-End

A diferencia del proceso de desarrollo del frontend el cual fue bastante directo y sencillo, el desarrollo del backend fue mucho más complejo debido a la experiencia limitada con las tecnologías empleadas en esta área del proceso. Esto significó una curva de aprendizaje mucho más empinada en comparación a la experimentada con la aplicación React. Particularmente el uso de máquinas EC2 y storage S3 de AWS, el primer acercamiento a dominios y DNS, y el uso de balanceadores de carga.

3.4.1. Etapa 0: Selección de tecnologías

Debido a la inexperiencia con aplicaciones web fuera de un entorno de desarrollo, siguiendo consejos de profesores y compañeros, se consideró utilizar un Elastic Beanstalk de AWS, sin embargo, la aplicación es muy liviana como para hacer uso de las capacidades de procesamiento y almacenamiento elástico que proporciona Elastic Beanstalk. Posteriormente se consideró brevemente el uso de un computador personal que pudiera tomar el rol de servidor para mantener la API REST viva. Sin embargo, la idea fue descartada rápidamente en pos de no aumentar el consumo eléctrico, y por ende la cuenta de la luz.

Luego se consideró utilizar VERCEL tanto para el frontend como para el backend, sin embargo, al tener la necesidad de guardar vídeos para poder procesarlos después, y que al tiempo de estar comparando opciones para el hosting de la API REST no se pudo encontrar información sobre el almacenamiento máximo que tienen las máquinas de vercel, se decidió no tomar esta opción.

Finalmente, se tomó la decisión de que la solución más básica probablemente sería la correcta en este caso y se optó por tener una máquina virtual EC2 de AWS. Además, a pesar de que esta elección representaría más trabajo al momento de configurar todo el trabajo de configuración necesario para que pudiera aceptar conexiones por internet provenientes de la página web hosteada en VERCEL, incurrió un proceso de aprendizaje fundamental para el entendimiento de cómo funcionan las conexiones de internet, y lo necesario para poder aceptar conexiones a distintos puertos, como funciona un DNS y los dominios asociados, estándares de conexión par navegadores modernos, entre otras cosas.

Amazon ofrece un gran rango de servicios de sus máquinas EC2 [[AWS, 2024a](#)]:

- Instancias de uso general
- Instancias optimizadas para computación
- Instancias optimizadas para memoria
- Instancias de computación acelerada
- Instancias de almacenamiento optimizado
- Instancias HPC (High Performance Computing)

Las cuales entregan una amplia gama de utilidades, como lo son múltiples opciones de almacenamiento, almacenamiento elástico en bloques, cluster networking, entre otros.

En base a que los requerimientos de la aplicación son, sin un orden de importancia particular:

1. Aceptar Conexiones
2. Guardar los videos momentáneamente
3. Poder procesar los videos
4. Ser capaz de servir una API REST en Node.js

No fue necesario crear una instancia EC2 especializada, por lo que se tomó la decisión de una máquina de uso general t2.micro la cual proporciona [\[AWS, 2024b\]](#):

Recurso	Descripción
VCPU	1
RAM	1 GB
Almacenamiento	Hasta 30 GB en la tier gratuita

Tabla 1: Especificaciones de la Máquina

Sin embargo, para el almacenamiento se optó por dejar 20 GB, ya que fue estimado como suficiente.

3.4.2. Primera Etapa: Levantamiento de Máquina

El levantamiento de máquina EC2 de AWS fue la primera gran barrera debido a la poca familiaridad con la plataforma y a la configuración general de máquinas virtuales. La configuración consistió de cinco pasos:

1. **Configuración de Máquina:** Debido a que el proceso era para una máquina normal sin almacenamiento extra ni fue necesario que se conectara a una VPN, el proceso de configuración se hizo utilizando todas las configuraciones estándares y por defecto de AWS exceptuando la elección de sistema operativo (se optó por Ubuntu Jammy V22.04).
2. **Apertura de puertos físicos:** Una vez creada la máquina, para que esta aceptara peticiones tipo HTTP y HTTPS, se debieron abrir los puertos 80 y 443 respectivamente. Además, las máquinas de Amazon vienen por defecto con el puerto 22 para conexiones tipo SSH abiertas.
3. **Habilitación de puertos en firewall:** Por seguridad, todos los puertos vienen cerrados en el firewall de las máquinas por lo que se debe cambiar las reglas de la misma para poder aceptar peticiones en los puertos correspondientes. En este caso se permitió que la máquina aceptara conexiones por los mismos puertos que se habían abierto anteriormente.
4. **Instalación de programas necesarios:** Para poder ejecutar todos los programas fue necesario instalar la última versión de Node.js en el momento: V18.19.0. Node.js además viene con el gestor de paquetes npm [npm, 2024b], el cual se instaló con la versión V10.2.3.
5. **Descarga de repositorio y paquetes:** Una vez que Node estuviera instalado en la máquina lo único que resta por hacer en este paso es descargar el repositorio desde Github e instalar los paquetes que el servidor necesita para funcionar. Esto se hace tomando ventaja del archivo package.json que se crea y actualiza automáticamente en los proyectos que utilizan Node [npm, 2024a].

Sin embargo, estas configuraciones no son suficientes para que la aplicación sea alcanzable por todas las conexiones procedentes de cualquier parte del mundo a pesar de que la máquina tiene una ip pública estática.

3.4.3. Conexión a Internet

Debido a que toda conexión hecha por internet fuera de una red de área local requiere de un dominio y que haya propagación de este por los servicios DNS para que llegue a los distintos ISP al rededor del mundo, se decidió utilizar el servicio de AWS Route 53 [AWS, 2024c] que permite la compra de dominios y gestión de redirecciones y subdominios del mismo. El dominio comprado fue memoria-nt.com.

Adicional a lo anterior, debido a que la API REST en Node.js escucha en el puerto 3001 y la conexión por internet debe ser por puerto 80 HTTP o 443 HTTPS, se decidió utilizar un balanceador de carga para poder hacer las redirecciones necesarias a los s correspondientes: toda conexión o petición realizada mediante protocolo HTTP o HTTPS al dominio memoria-nt.com fue redireccionada al balanceador de carga, el cual envía las peticiones al puerto 3001 de la máquina con IP pública estática **3.144.135.16**. Sin embargo, los balanceadores de carga no son baratos, lo que implicó que el precio del mismo fuera mucho más elevado que el precio de la máquina, por lo que durante el resto del tiempo del desarrollo del proyecto se buscó una manera de reemplazarlo.

Posteriormente el balanceador de carga fue reemplazado con una redirección interna utilizando un reverse proxy con NGINX [NGINX, 2024], el cual cumple la misma función que el balanceador de carga de enrutar todas las peticiones HTTP o HTTPS hacia el puerto 3001, pero libre de costo. El único cambio de configuraciones necesario fue el de cambiar las redirecciones tipo A desde Route 53 para que, en vez de apuntar al balanceador de carga, apuntan directo a la máquina y las enrutan al puerto correspondiente.

Adicional a esto, se debieron configurar los security groups de la máquina EC2 con las inbound rules mostradas en la figura 7. Estas actúan como un cortafuego o firewall a nivel de red, por lo que son necesarias si se desea aceptar conexiones entrantes a la máquina como lo serían llamados a los endpoints de la API.

Security group rule...	IP version	Type	Protocol	Port range	Source	Description
sgr-0dc469c04f4faaf55	IPv4	Custom TCP	TCP	3001	0.0.0.0/0	Puerto api
sgr-05a191218066a0...	IPv4	HTTPS	TCP	443	0.0.0.0/0	puerto https
sgr-01c20f3064df958d1	IPv4	HTTP	TCP	80	0.0.0.0/0	puerto http
sgr-0ada7975b16a8940	IPv4	SSH	TCP	22	0.0.0.0/0	-

Figura 7: : Inbound rules de Máquina EC2

Fuente: Elaboración propia.

Y con esta configuración finalmente es posible que las peticiones del front-end puedan alcanzar a la api, y que esta pueda devolver las respuestas correspondientes.

3.5. Nuevo Modelo

El funcionamiento del modelo anterior puede ser visto en el apartado de anexos. La información presentada es importante para contextualizar la sección actual.

Para comenzar la implementación del nuevo modelo se utilizaron algunos módulos del modelo anterior como base, particularmente el driver de video, el reconocimiento de caras y el extractor de características.

Estos módulos del modelo anterior fueron utilizados para transformar los videos en datos más sencillos de procesar: datos JSON con la posición de todos los puntos del facemesh en cada frame más la certeza de gestos realizados en cada frame entregado por mediapipe.

3.6. Recolección de datos

Como fue mencionado anteriormente los datos son almacenados tanto en un csv (datos personales de los usuarios y pacientes), en los archivos directos (videos de los gestos), y los archivos JSON que fueron obtenidos como resultado del preprocesador creado en la sección anterior. Sin embargo, como se puede ver en la sección [3.4.1](#), la instancia t.2 micro solamente puede tener hasta 30 GB de almacenamiento, lo que presenta una limitante si la cantidad de usuarios aumenta con el tiempo. Este problema fue enfrentado utilizando otro de los servicios entregados por AWS: Amazon Simple Storage Service o también conocido con el acrónimo de S3. S3 es un sistema de almacenamiento de objetos en la nube que, al igual que el almacenamiento de un computador común y corriente, permite almacenar todo tipo de archivos como los son archivos de texto, documentos, planillas, presentaciones, audio, imágenes y video, entre otros. Además, al ser un servicio entregado por AWS, transferir archivos desde y hacia S3 dentro del ecosistema de Amazon es muy sencillo. [\[Services, 2024\]](#)

Fueron creados dos buckets, que son la unidad de almacenamiento de S3 (tal como cada unidad de computación de AWS EC2 es una instancia), las cuales cumplen dos fines distintos pero esenciales para el crecimiento de la base de datos en el tiempo. La primera instancia que fue creada cumple el rol de tener un lugar en donde poder almacenar los videos directamente en caso de necesitar reentrenar el modelo y también para poder ir de manera periódica liberando el espacio de almacenamiento usado de la máquina EC2. El segundo bucket que fue creado cumple la función de tener un lugar centralizado en donde poder almacenar tanto los datos en JSON, los consentimientos informados de los usuarios, y el archivo csv que almacena los datos personales de los usuarios y pacientes.

Los datos fueron en su gran mayoría recolectados gracias a la amabilidad de tanto profesores como estudiantes de la Facultad de Medicina Campus Norte de la Universidad de Chile, además de conocidos y familiares de los miembros del equipo de investigación.

3.7. Análisis experimental de modelos de aprendizaje automático

Una vez que los datos fueron obtenidos surge la interrogante de que modelo de aprendizaje automático utilizar, sin embargo, antes de poder alimentar cualquier modelo de ML, es recomendado preparar los datos de antemano. Esta preparación de datos comprende un proceso de múltiples pasos entre los cuales se encuentran [García *et al.*, 2016]:

1. Limpiar la data
2. Normalizar la data
3. Particionar la data
4. Transformar la data

Afortunadamente debido a todo el trabajo realizado hasta el momento, algunos puntos ya han sido solucionados. Primero, los datos están limpios desde la toma de los datos debido a que para cada gesto el video se capturó individualmente, eliminando así el ruido y dejando solamente el gesto. Segundo, los datos fueron transformados utilizando mediapipe mediante la implementación de los módulos del modelo anterior. Esto también cumplió la función de normalización de datos ya que la probabilidad entregada de cada gesto en la lista es un número decimal entre cero y uno.

3.7.1. Granularidad

Además de la limpieza de datos brindada por la plataforma de captura de videos, otro paso lógico que se puede tomar en la limpieza de videos es la granularización de los cuadros por segundo (definiendo granularidad como la distancia entre un punto de captura y el siguiente, lo que determina la capacidad de apreciar cambios entre ellos), lo que ya fue enfrentado en parte al limitar el framerate de los videos. Sin embargo, aún a 30 cuadros por segundo la diferencia entre un cuadro y el siguiente es pequeña o imperceptible. Es por esto que se decidió disminuir la granularidad temporal de los datos.

3.7.2. Reducción de dimensionalidad

Debido a que la lista de expresiones que detecta mediapipe es bastante extensa (ver tabla 16) se decidió implementar una técnica común en el análisis de datos: la reducción de dimensionalidad. La implementación de la reducción de dimensionalidad simplifica el conjunto de datos para conservar la información más importante, lo que ayuda a reducir la complejidad computacional y reducir el ruido al eliminar aquellos componentes que no aportan datos

importantes en la clasificación. Además al tener una cantidad acotada de datos la implementación de los clasificadores será más sencilla ya se podrá evitar el sobre ajuste. Más aún, se podrán utilizar las métricas para poder saber cuáles son las características más importantes, lo que puede ayudar a mejorar el modelo en una implementación futura.

Las técnicas de reducción de dimensionalidad elegidas fueron PCA y SVD debido a la familiaridad con ambas. PCA o Análisis de Componentes Principales por sus siglas en inglés y SVD o Descomposición de Valores Singulares por sus siglas en inglés

La razón tras la implementación de dos métodos de reducción de dimensionalidad es poder generar una comparación y poder cuantificar las diferencias entre la implementación de las combinaciones de modelos con algoritmos de reducción de dimensionalidad para poder así generar el mejor clasificador posible en base a los conocimientos del equipo de investigación.

3.7.3. Regresor

Luego de realizar una reducción de dimensionalidad, se plantea el uso de regresores para poder transformar las matrices resultantes en datos que puedan ser consumidos por un clasificador. Este paso fue realizado debido a que a pesar de obtener un conjunto de datos que aportan la mayor parte de información para detectar los gestos faciales, estos modelos de reducción de dimensionalidad no son el mejor acercamiento si es que se desea encontrar relaciones complejas entre las características a estudiar. [Brenning, 2023]

3.8. Clasificador

Para finalizar, se crea un pipeline con los pasos mencionados anteriormente para poder dejar el regresor pre entrenado con el fin de expedir el preprocesamiento de los datos antes de entregárselos al clasificador, que en este caso es SVM o Support Vector Machine.

En total el proceso de entrenamiento del clasificador consiste en los siguientes pasos:

1. Captura de datos.
2. Transformación de videos.
3. Unión de los datos considerando cantidades de frames.
4. Reducción de dimensionalidad de los datos.
5. Separación de conjuntos de entrenamiento y pruebas.
6. Entrenamiento de regresor.

7. Creación de pipeline.
8. Entrenamiento de clasificador.
9. Exportación de clasificador preentrenado.

CAPÍTULO 4

VALIDACIÓN DE LA SOLUCIÓN

El punto inicial de la validación, posterior a la recolección de datos y videos, fue la transformación de los mismos a una cadena de información en formato JSON donde cada elemento del documento corresponde a la información del face mesh y la probabilidad de expresiones para cada frame de un video. Estos se guardan con el nombre del gesto correspondiente, en una carpeta con el mismo nombre hash que se genera cuando se guardan los videos en una primera instancia. Posteriormente, se cargan los datos de cada uno de los archivos JSON y se dividen en tres para cada archivo, para poder tomar segundo tercio ya que así se puede asegurar que los frames en donde efectivamente está siendo realizado el movimiento queden registrados con el menor ruido posible, o en otras palabras, debido a que todos los videos tienen un breve periodo de tiempo al inicio y al final en donde las personas reaccionan antes de realizar el gesto y presionan el botón para terminar la grabación, al tomar solamente los frames del medio de la secuencia se asegura captar los gestos sin ese periodo de tiempo al principio y al fin. A continuación, fueron extraídos los datos de las expresiones y concatenados en un dataframe que tiene en las columnas los nombres de las expresiones y en las filas la información de cada frame. La estructura del dataframe puede verse en la tabla 3.8, donde la columna **file** corresponde al nombre del archivo, **frame_index** corresponde al número de frame en el video (donde el frame 0 es el primero), y las columnas **exp_1**, **exp_2** y **exp_3** corresponden a la probabilidad de la expresión entregada por MediaPipe.

file	frame_index	exp_1	exp_2	exp_3	...
AbrirYCerrarBoca.json	0	x_1	y_1	z_1	...
AbrirYCerrarBoca.json	1	x_2	y_2	z_2	...
AbrirYCerrarBoca.json	2	x_3	y_3	z_3	...
⋮	⋮	⋮	⋮	⋮	⋮

Tabla 2: Estructura general del dataframe de los datos.

Fuente: Elaboración propia

Luego de generar el dataframe se puede continuar con el preprocesamiento de los datos para poder entregárselo a los distintos regresores, el cual inicia con el aumento de la granularidad temporal de los datos. Esto consistió en seleccionar uno de cada n frames para los datos. En donde se eligieron $n = 1$, $n = 2$, $n = 3$ y $n = 5$. Además, la elección de si el número es par o impar permite la selección automática del frame medio, en donde se sabe que se estará realizando el movimiento con 100 % de certeza.

Posteriormente se realizó tanto una reducción de dimensionalidad con PCA como con SVD, en donde se tomaron $n = 2$, $n = 3$, $n = 5$ y $n = 7$ componentes principales/vectores singulares. La cantidad de componentes, sus varianzas y relaciones fueron graficadas y comparadas para obtener el mejor posible resultado tomando en cuenta rendimiento y

tiempo de ejecución.

Finalmente los resultados de las combinaciones de cantidad de frames por segundo y cantidad de componentes fueron separados en conjuntos de entrenamiento y prueba de manera aleatoria utilizando la función `train_test_split` provista por la biblioteca `scikitlearn` [Scikit-learn Developers, 2024a] en una razón de 80 - 20 entre los conjuntos de entrenamiento y prueba utilizando la misma semilla para que los datos puedan ser reproducibles, para luego ser procesados por los tres distintos regresores propuesto y evaluados utilizando los mismos conjuntos de entrenamiento y prueba para llegar a un gráfico que muestra el error cuadrático medio para los tres regresores vs sus desempeños con PCA y con SVD.

En las siguientes secciones serán explorados los rendimientos de cada una de las variables

1. **Cantidad de frames = F**
2. **Cantidad de componentes = C**
3. **Regresor utilizado**

En donde el rendimiento final será entendido como el error cuadrático medio (o **MSE** por sus siglas en ingles) de una determinada combinación de F , C y regresor.

4.1. Cantidad de frames

En esta sección serán explorados los **MSE** totales de las distintas combinaciones con los F propuestos en la sección anterior, además del promedio y varianza de los mismos agrupados por técnica de reducción de dimensionalidad y del conjunto en su totalidad. Para concluir será mostrado un resumen de todos los datos de interés.

Para poder comparar el **MSE** de la cantidad de frames se decidió (sin ningún particular) el uso de cinco componentes principales para poder comparar así el **MSE**.

El resultado del regresor con $F = 5$ puede ser visto en la figura 8 y la tabla 3. El mejor resultado con un **MSE** de $1,942e^{-3}$ corresponde a XGB + PCA, mientras que el peor resultado con un **MSE** de $2,241e^{-3}$ corresponde a GB + SVD. Además, por técnica de reducción de dimensionalidad fueron obtenidos los promedios para PCA= $2,001e^{-3}$ y para SVD $2,148e^{-3}$, y las varianzas para PCA= $2,842536e^{-9}$ y para SVD = $2,618652e^{-9}$, lo que entrega un **MSE** promedio total de $2,0925366616255776e^{-3}$ y una varianza promedio total de $1,2161150947240095e^{-8}$.

Modelo	MSE
XGBoost - PCA	0.001942
Gradient Boosting - PCA	0.002016
Random Forest - PCA	0.002046
Random Forest - SVD	0.002143
XGBoost - SVD	0.002168
Gradient Boosting - SVD	0.002241

Tabla 3: Resultados de Modelos para $F = 5$
Fuente: Elaboración propia.

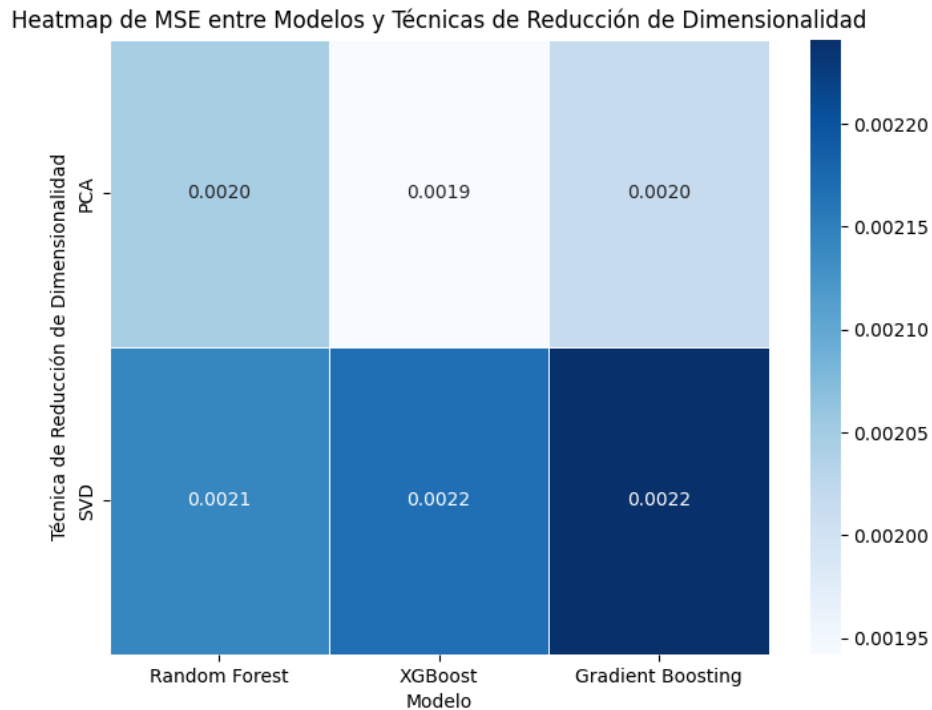


Figura 8: Error cuadrático medio para los datos con $F = 5$.
Fuente: Elaboración propia

El resultado del regresor con $F = 3$ puede ser visto en la figura 9 y la tabla 4. El mejor resultado con un **MSE** de $1,223e^{-3}$ corresponde a RF + SVD, mientras que el peor resultado con un **MSE** de $1,661e^{-3}$ corresponde a GB + SVD. Además, por técnica de reducción de dimensionalidad fueron obtenidos los promedios para PCA= $1,353e^{-3}$ y para SVD $1,477e^{-3}$, y las varianzas para PCA= $1,189601e^{-8}$ y para SVD = $5,147384e^{-8}$, lo que entrega un **MSE** promedio total de $1,415030055298156e^{-3}$ y una varianza promedio total de $2,996246489506508e^{-8}$.

Model	MSE
Random Forest - SVD	0.001223
XGBoost - PCA	0.001289
Random Forest - PCA	0.001291
Gradient Boosting - PCA	0.001479
XGBoost - SVD	0.001547
Gradient Boosting - SVD	0.001661

Tabla 4: Resultados de Modelos para $F = 3$
Fuente: Elaboración propia.

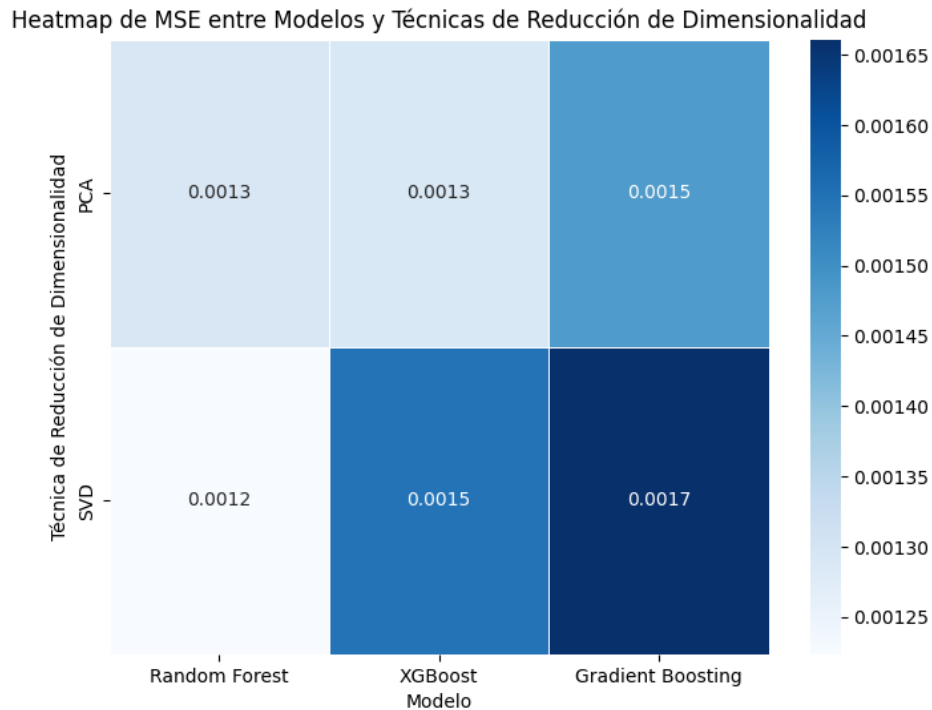


Figura 9: Error cuadrático medio para los datos con $F = 3$.
Fuente: Elaboración propia

El resultado del regresor con $F = 2$ puede ser visto en la figura 10 y la tabla 5. El mejor resultado con un MSE de $1,122e^{-3}$ corresponde a XGB + PCA, mientras que el peor resultado con un MSE de $1,637e^{-3}$ corresponde a GB + SVD. Además, por técnica de reducción de dimensionalidad fueron obtenidos los promedios para PCA= $1,266e^{-3}$ y para SVD $1,349e^{-3}$, y las varianzas para PCA= $3,231652e^{-8}$ y para SVD = $6,2901964e^{-8}$, lo que entrega un MSE promedio total de $1,3075393444691233e^{-3}$ y una varianza promedio total de $4,017171413814802e^{-8}$.

Model	MSE
XGBoost - PCA	0.001122
Random Forest - SVD	0.001178
Random Forest - PCA	0.001208
XGBoost - SVD	0.001233
Gradient Boosting - PCA	0.001467
Gradient Boosting - SVD	0.001637

Tabla 5: Resultados de Modelos para $F = 2$
Fuente: Elaboración propia.

Heatmap de MSE entre Modelos y Técnicas de Reducción de Dimensionalidad

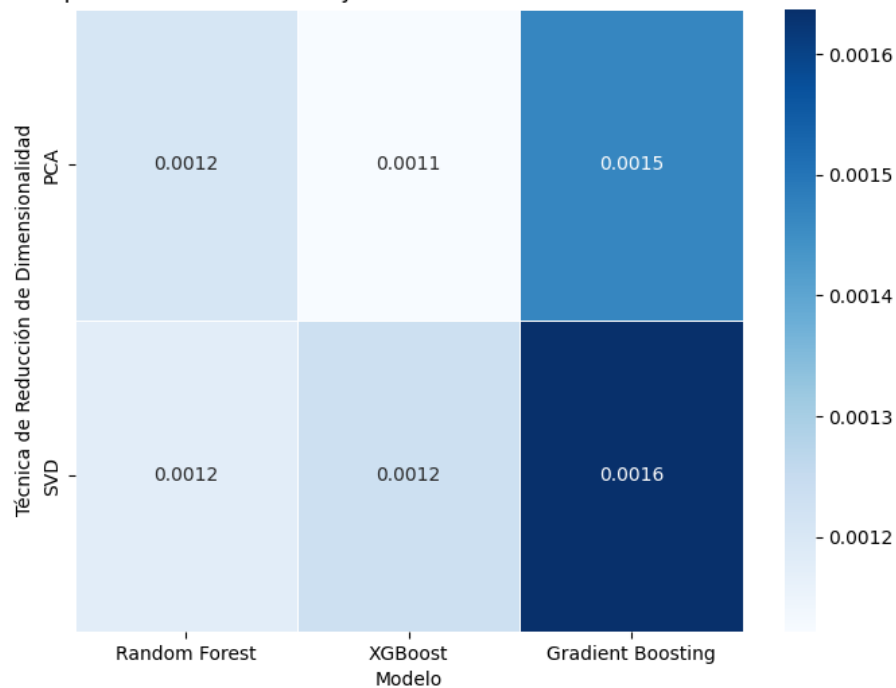


Figura 10: Error cuadrático medio para los datos con $F = 2$.
Fuente: Elaboración propia

El resultado del regresor con $F = 1$ puede ser visto en la figura 11 y la tabla 6. El mejor resultado con un **MSE** de $7,38e^{-4}$ corresponde a RF + SVD, mientras que el peor resultado con un **MSE** de $1,35e^{-3}$ corresponde a GB + SVD. Además, por técnica de reducción de dimensionalidad fueron obtenidos los promedios para PCA= $9,97e^{-4}$ y para SVD $9,96e^{-4}$, y las varianzas para PCA= $6,612442e^{-8}$ y para SVD = $1,00355e^{-7}$, lo que entrega un **MSE** promedio total de $1,3075393444691233e^{-3}$ y una varianza promedio total de $4,017171413814802e^{-8}$.

Model	MSE
Random Forest - SVD	0.000738
Random Forest - PCA	0.000846
XGBoost - PCA	0.000851
XGBoost - SVD	0.000901
Gradient Boosting - PCA	0.001294
Gradient Boosting - SVD	0.001350

Tabla 6: Error cuadrático medio para los datos con $F = 1$.
Fuente: Elaboración propia

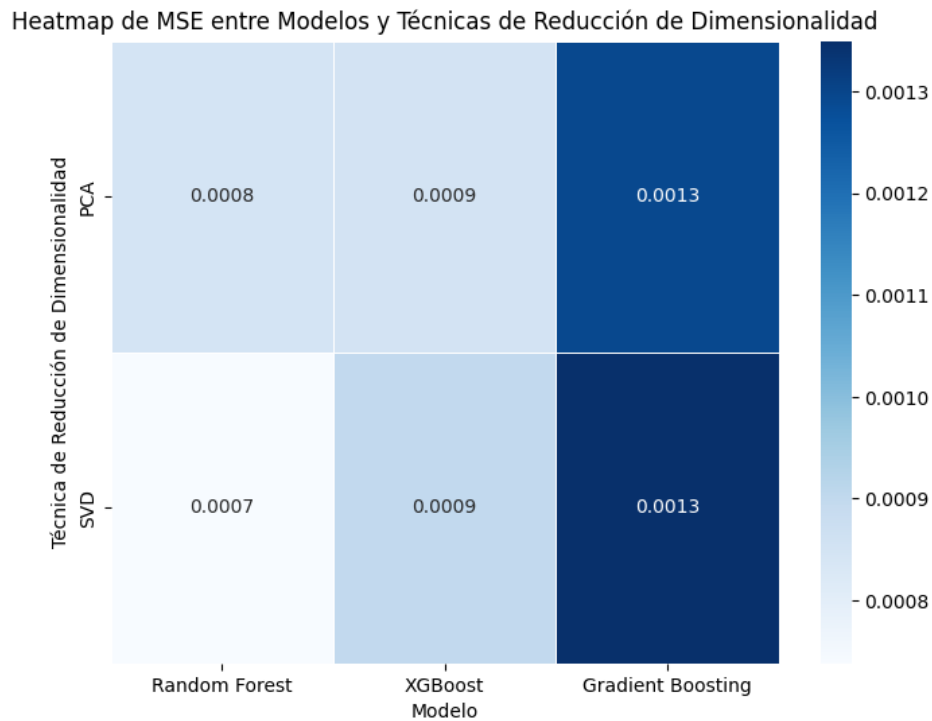


Figura 11: Error cuadrático medio para datos con $F = 1$.
Fuente: Elaboración propia

Para graficar como se relaciona el **MSE** con F , fueron considerados los **MSE** de RF, ya que esta fue el regresor que mostró el mejor resultado final.

Al comparar los resultados se pueden observar los siguientes comportamientos.

- Existe una clara proporcionalidad inversa entre la cantidad de frames considerados para el regresor versus el rendimiento de este.
- Entre ambas técnicas de reducción de dimensionalidad no se aprecia que una entregue constantemente resultados mejores.

- Al comparar el rendimiento de los distintos regresores podemos aseverar que si existe una diferencia considerable entre estos ya que constantemente Gradient Boosting tiene el peor rendimiento promedio.
- A pesar de que los valores de **MSE** si se encuentran relacionados a F , y por lo tanto el promedio del mismo también, la varianza de las mismas no presentan una relación con el **MSE** por lo que no entregan información relevante.
- El mejor resultado fue para $F = 1$, esto significa que, al menos a 30 FPS, la diferencia entre cada frame es la suficiente para no sobrepoblar los datos con entradas muy similares. Sin embargo, es importante volver a destacar que los videos fueron capturados con un límite de 30 FPS, lo que implica una diferencia mayor entre cada dos cuadros consecutivos que si el video tuviera 60 FPS como es el estándar actual.

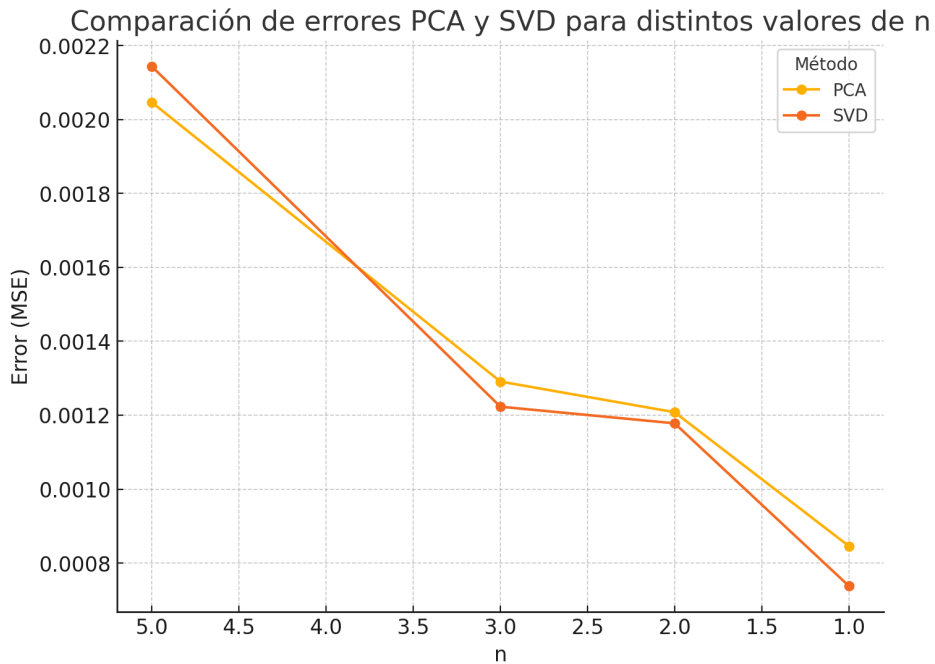


Figura 12: Comparación de **MSE** con F para técnicas de reducción de dimensionalidad
Fuente: Elaboración propia

4.2. Cantidad de componentes

A diferencia del punto anterior en donde se tomó un C arbitrario para comparar el impacto de F en el error cuadrático medio del regresor, debido al resultado obtenido en la sección anterior se seleccionó $F = 1$ para asegurar que los resultados sean lo más certeros posibles. En esta sección además se selecciona el mismo C para PCA y SVD con el fin de agilizar el proceso de análisis.

Los resultados entregados para $C = 2$ pueden ser vistos en la tabla 7 y figura 13. EL MSE menor es de $5,108e^{-3}$ para RF + PCA con un promedio de $5,373512611625434e^{-3}$.

Model	MSE
Random Forest - PCA	0.005108
Random Forest - SVD	0.005203
Gradient Boosting - PCA	0.005315
Gradient Boosting - SVD	0.005413
XGBoost - PCA	0.005531
XGBoost - SVD	0.005671

Tabla 7: Error cuadrático medio para los datos con $C = 2$.
Fuente: Elaboración propia

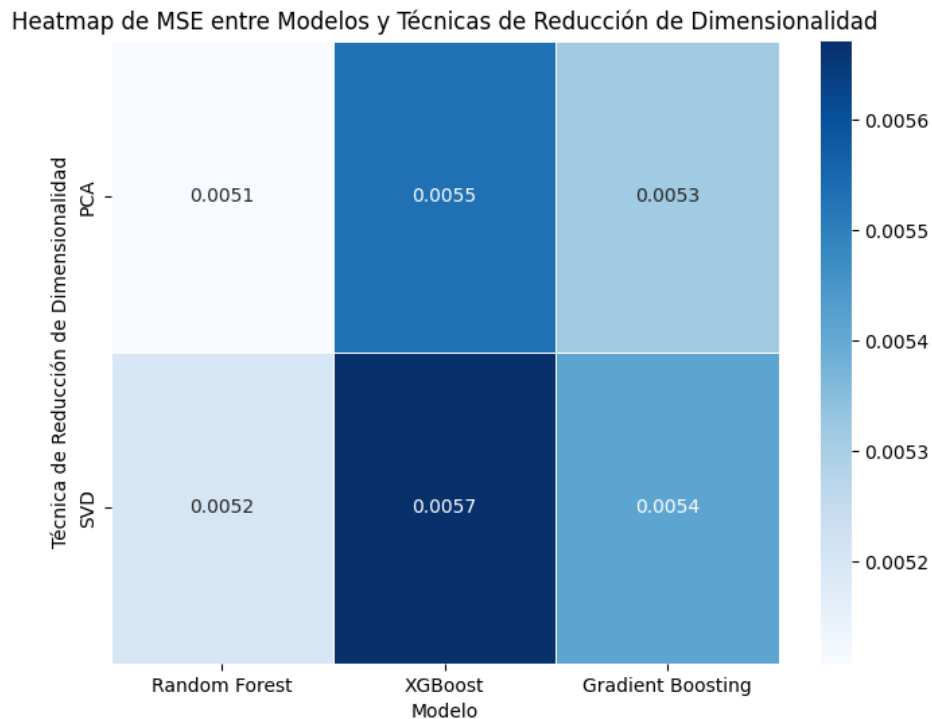


Figura 13: Error cuadrático medio para datos con $C = 2$.
Fuente: Elaboración propia

Adicionalmente, puede ser observada la varianza explicada de los datos en la figura 14 para PCA, en donde dos componentes solamente logran capturar el 47% de la variabilidad del conjunto original de datos. Por otra parte, la influencia de los componentes SVD sobre el conjunto de los datos puede ser visto en la figura 15, en donde puede ser apreciado que el primer componente aporta más del quíntuple de información sobre los datos. Como es de esperarse con una cantidad tan baja de Componentes para SVD y PCA el **MSE** del modelo es sumamente alto, más alto incluso que el peor resultado obtenido con $C = 5$ y $F = 5$, el cual fue el peor caso estudiado en la sección anterior. De momento los resultados son comparables entre ambos métodos de reducción de dimensionalidad, sin embargo, PCA presenta un mejor rendimiento que SVD con todos los modelos.

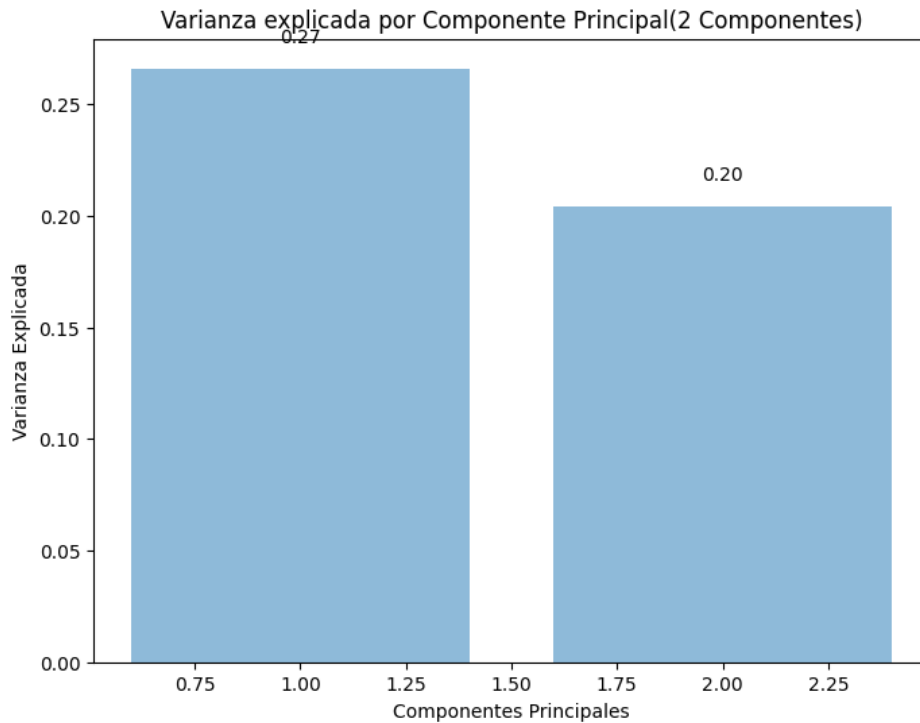


Figura 14: Varianza explicada de PCA con $C = 2$
Fuente: Elaboración propia

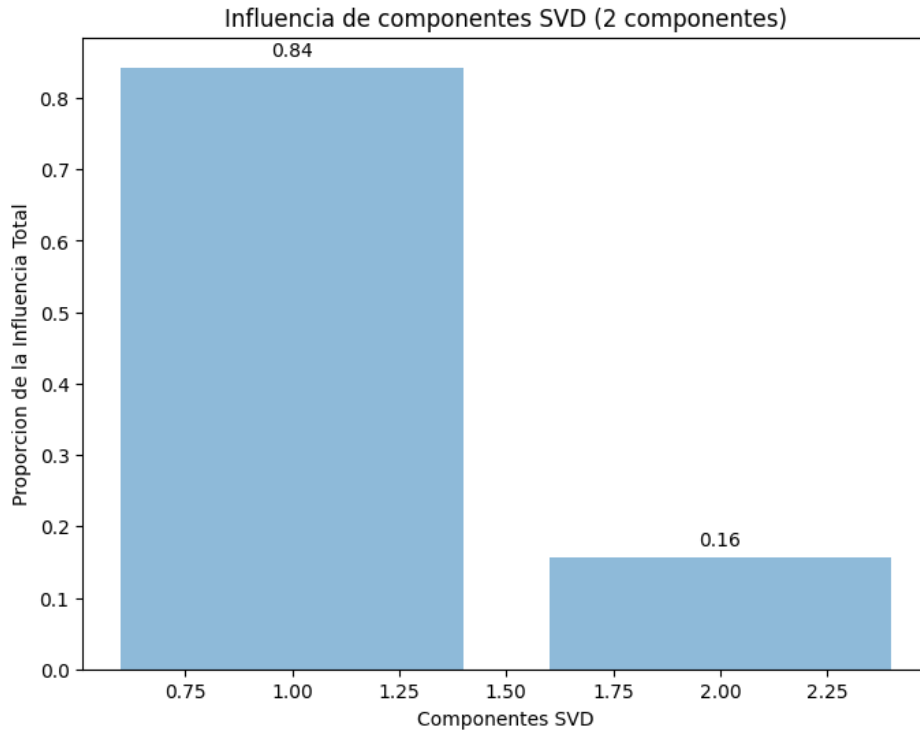


Figura 15: Influencia de los componentes SVD con $C = 2$
Fuente: Elaboración propia

Los resultados entregados para $C = 3$ pueden ser vistos en la tabla 8 y figura 16. EL MSE menor es de $2,035e^{-3}$ para RF + PCA nuevamente, con un promedio de $2,546368753510094e^{-3}$.

Model	MSE
Random Forest - PCA	0.002035
Random Forest - SVD	0.002080
XGBoost - PCA	0.002452
XGBoost - SVD	0.002509
Gradient Boosting - PCA	0.003081
Gradient Boosting - SVD	0.003123

Tabla 8: Error cuadrático medio para los datos con $C = 3$.
Fuente: Elaboración propia

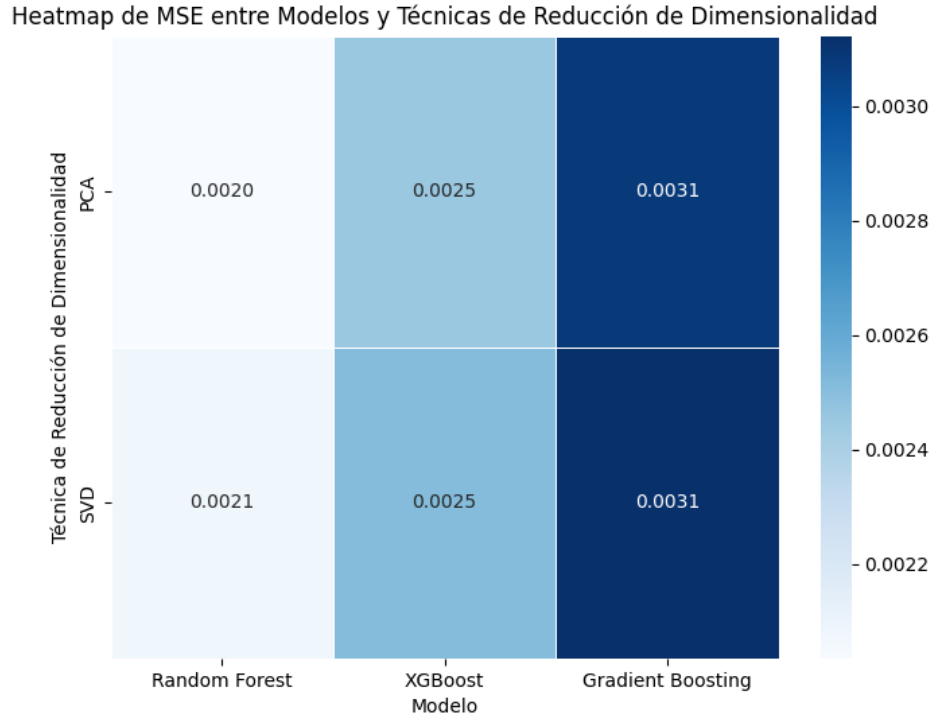


Figura 16: Error cuadrático medio para datos con $C = 2$.
Fuente: Elaboración propia

Adicionalmente, puede ser observada la varianza explicada de los datos en la figura 17 para PCA, en donde tres componentes logran capturar el 63% de la variabilidad del conjunto original de datos. Por otra parte, la influencia de los componentes SVD sobre el conjunto de los datos puede ser visto en la figura 18, en donde el primer componente ha disminuido la información que aporta a los datos, sin embargo, el segundo componente también ha disminuido, y el tercer componente aporta menos de un 10%. Desde ya se puede apreciar que la relación entre C y **MSE** es mucho estrecha que la relación entre F y **MSE** ya que al aumentar de dos a tres componentes el **MSE** se redujo a la mita en el peor caso (GB) y casi al triple en el mejor (RF). Además, PCA ya no presenta una ventaja significativa por sobre SVD, sino que el comportamiento de ambos se volvió bastante similar.

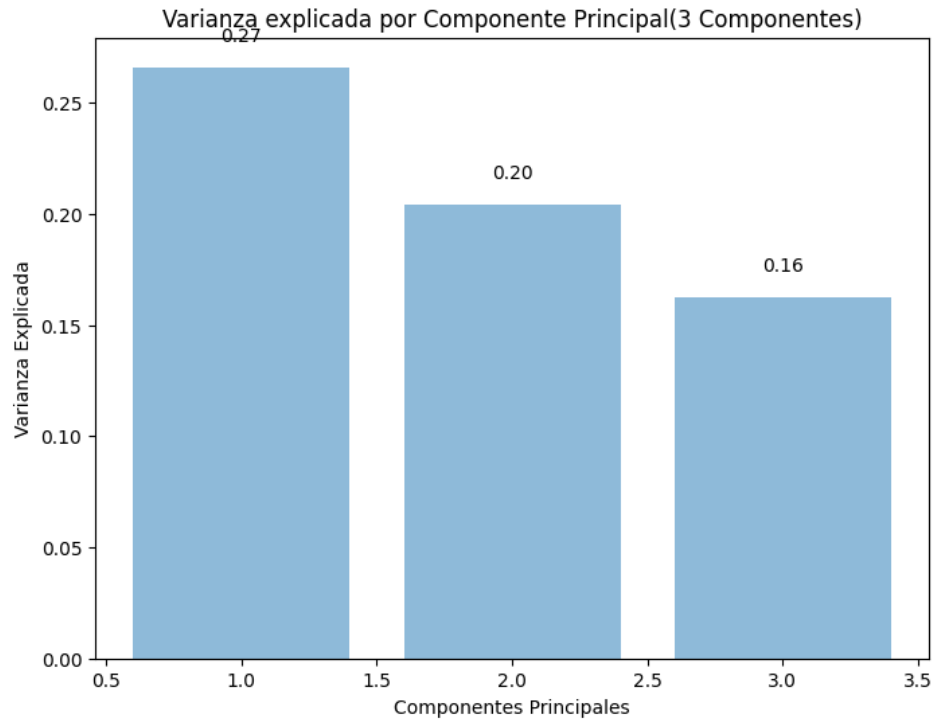


Figura 17: Varianza explicada de PCA con $C = 3$
Fuente: Elaboración propia

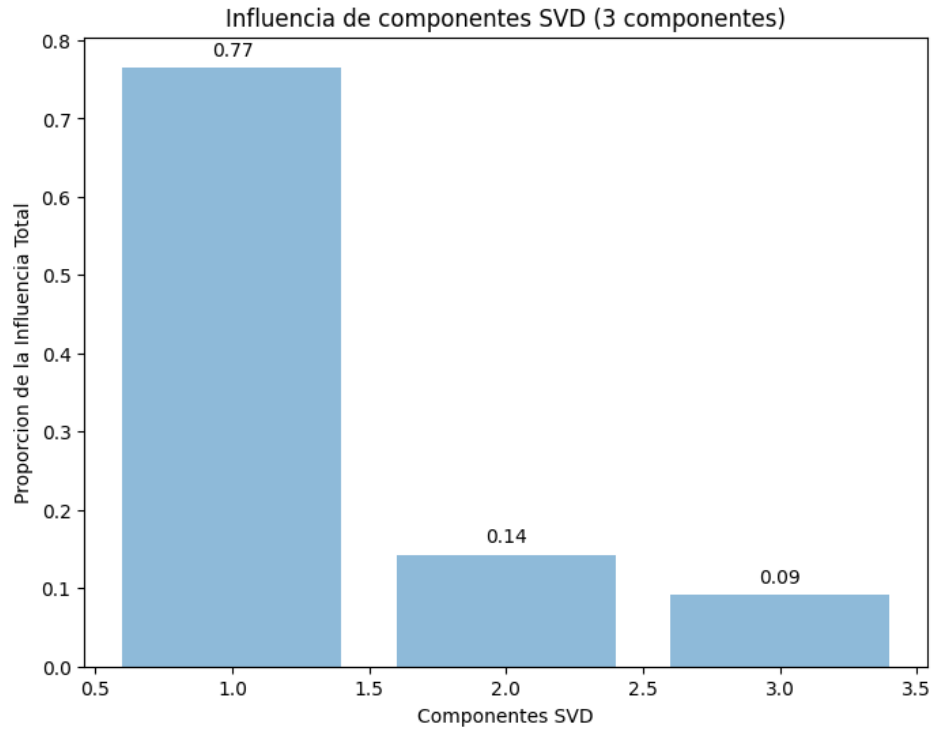


Figura 18: Influencia de los componentes SVD con $C = 3$
Fuente: Elaboración propia

Los resultados entregados para $C = 5$ pueden ser vistos en la tabla 6 y figura 8. Debido a que esta fue la primera combinación explorada en la sección anterior, los datos de **MSE** y promedios son los mismos.

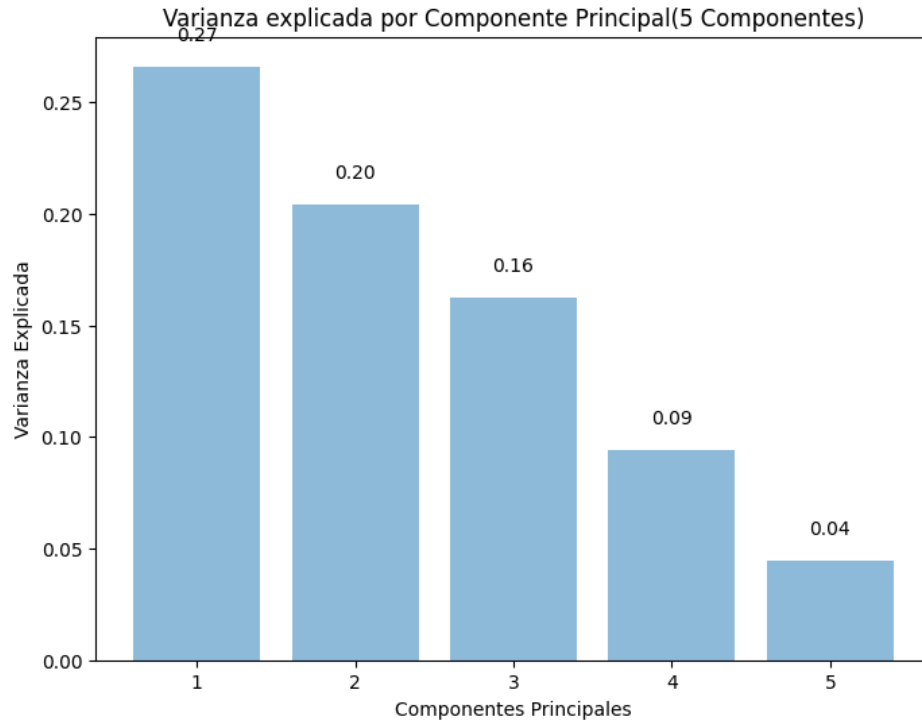


Figura 19: Varianza explicada de PCA con $C = 5$
Fuente: Elaboración propia

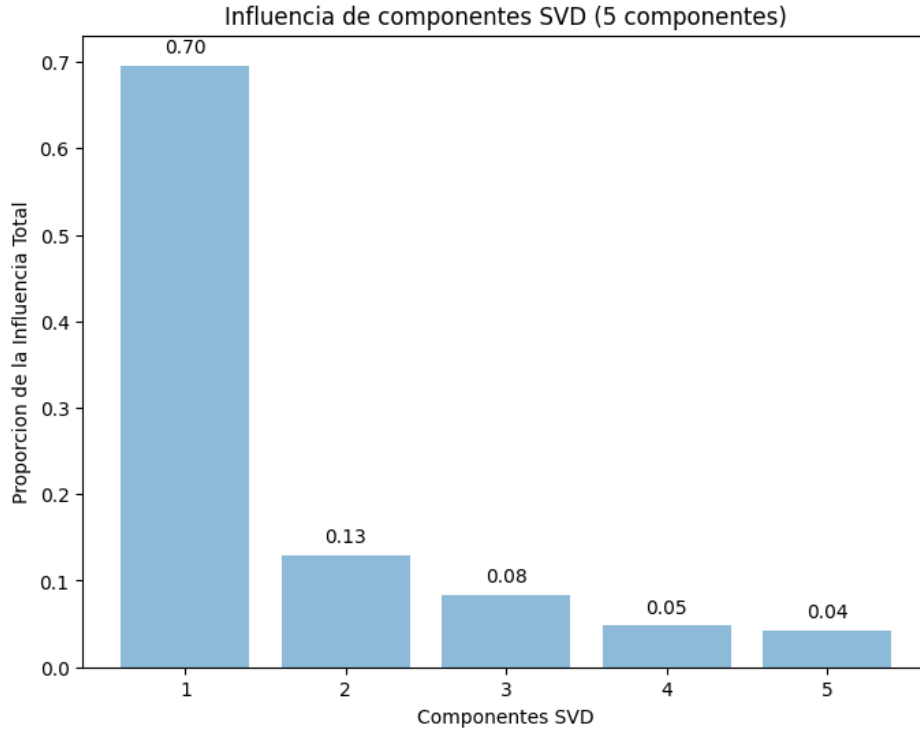


Figura 20: Influencia de los componentes SVD con $C = 5$
Fuente: Elaboración propia

Adicionalmente, puede ser observada la varianza explicada de los datos en la figura 19 para PCA, en donde cinco componentes logran capturar el 77 % de la variabilidad del conjunto original de datos. Por otra parte, la influencia de los componentes SVD sobre el conjunto de los datos puede ser visto en la figura 20, en donde el primer componente sigue aportando la gran mayoría de la información de los datos por sobre todos los otros componentes. El **MSE** promedio en este punto es bastante bueno, siendo el más bajo al que se puede llegar solamente disminuyendo el F para $C = 5$. Sin embargo, la diferencia entre los **MSE** de $C = 2$ y $C = 3$ es proporcionalmente mucho mayor que la diferencia entre $C = 3$ y $C = 5$, lo que sugiere un comportamiento de rendimientos decrecientes a medida que aumenta C , comportamiento que será reforzado con el análisis de $C = 7$.

Los resultados entregados para $C = 7$ pueden ser vistos en la tabla 9 y figura 21. EL **MSE** menor es de $4,75e^{-4}$ para XGB + PCA (aunque XGB + SVD tiene un **MSE** muy cercano con $4,81e^{-4}$), con un promedio de $6,09237510725375e^{-4}$.

Model	MSE
XGBoost - PCA	0.000475
XGBoost - SVD	0.000481
Random Forest - SVD	0.000544
Random Forest - PCA	0.000599
Gradient Boosting - PCA	0.000770
Gradient Boosting - SVD	0.000786

Tabla 9: Error cuadrático medio para los datos con $C = 7$.
Fuente: Elaboración propia

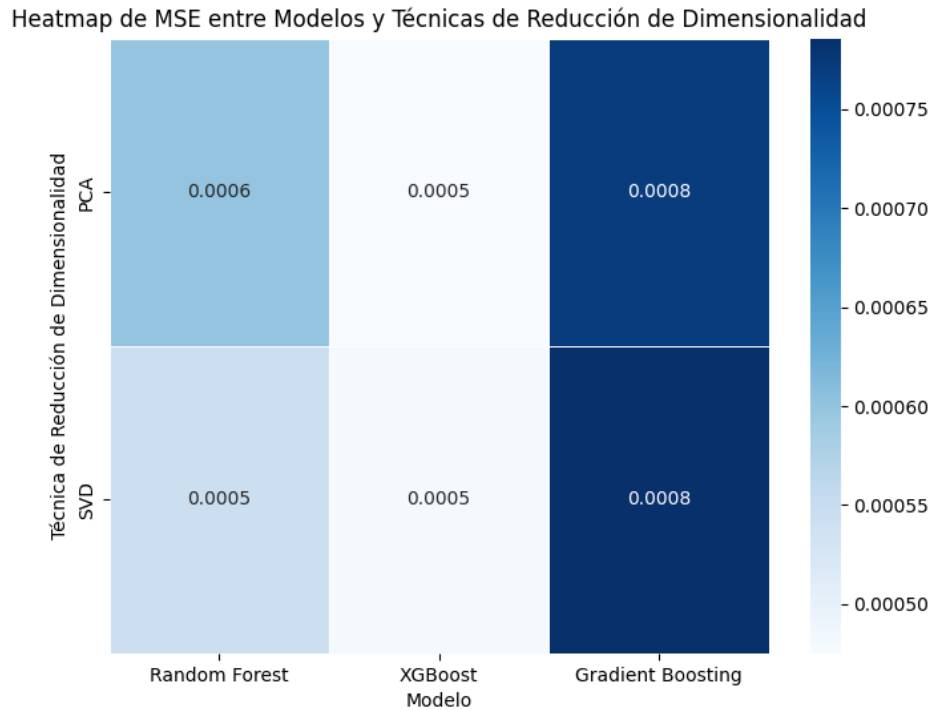


Figura 21: Error cuadrático medio para datos con $C = 2$.
Fuente: Elaboración propia

Adicionalmente, puede ser observada la varianza explicada de los datos en la figura 22 para PCA, en donde siete componentes logran capturar el 84% de la variabilidad del conjunto original de datos. Por otra parte, la influencia de los componentes SVD sobre el conjunto de los datos puede ser visto en la figura 23.

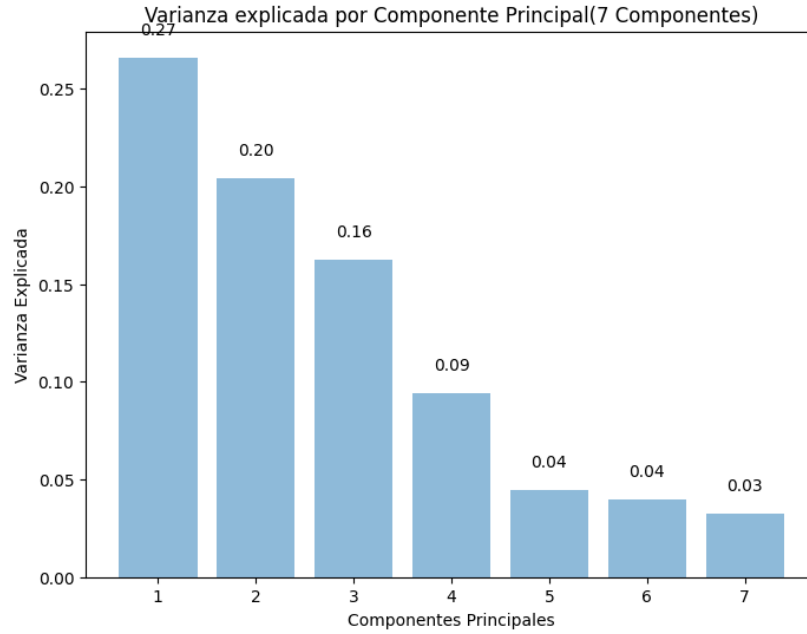


Figura 22: Varianza explicada de PCA con $C = 7$
Fuente: Elaboración propia

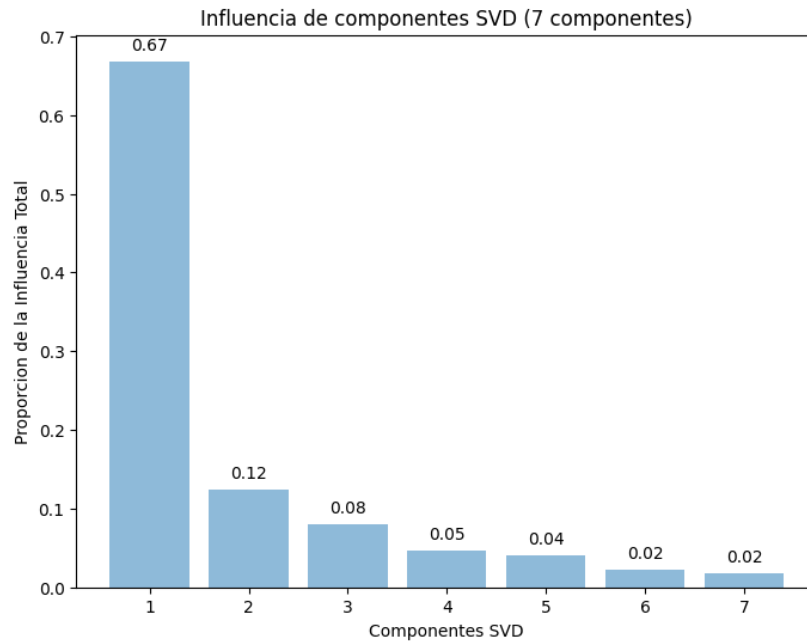


Figura 23: Influencia de los componentes SVD con $C = 7$
Fuente: Elaboración propia

Con $C = 7$ fue conseguido un **MSE** incluso menor que el mejor obtenido en el análisis de F ,

sin embargo, como fue planteado en el punto anterior, la mejora entre $C = 5$ y $C = 7$ no es muy significativa. Esto significa que, a pesar de presentar un comportamiento que claramente muestra una proporcionalidad entre C y el **MSE** del modelo, esta relación si se encuentra delimitada por rendimientos decrecientes. Sumando más a este punto importante, el tiempo de ejecución del código aumenta de manera considerable al crecer C como puede ser visto en la figura 24 para RF y XGB, y figura 25. La razón detrás de la separación de los gráficos fue que el entrenamiento de GB tomó consistentemente mucho más tiempo que el entrenamiento de XGB y RF.

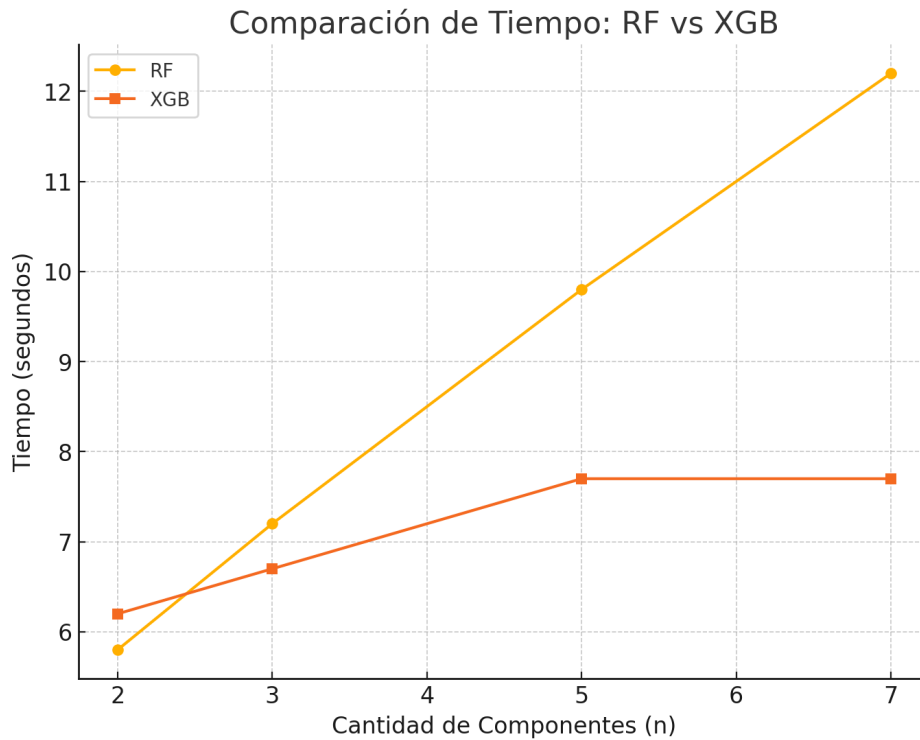


Figura 24: Grafico de Cantidad de Componentes vs tiempo que demoró entrenar el modelo
Fuente: Elaboración propia

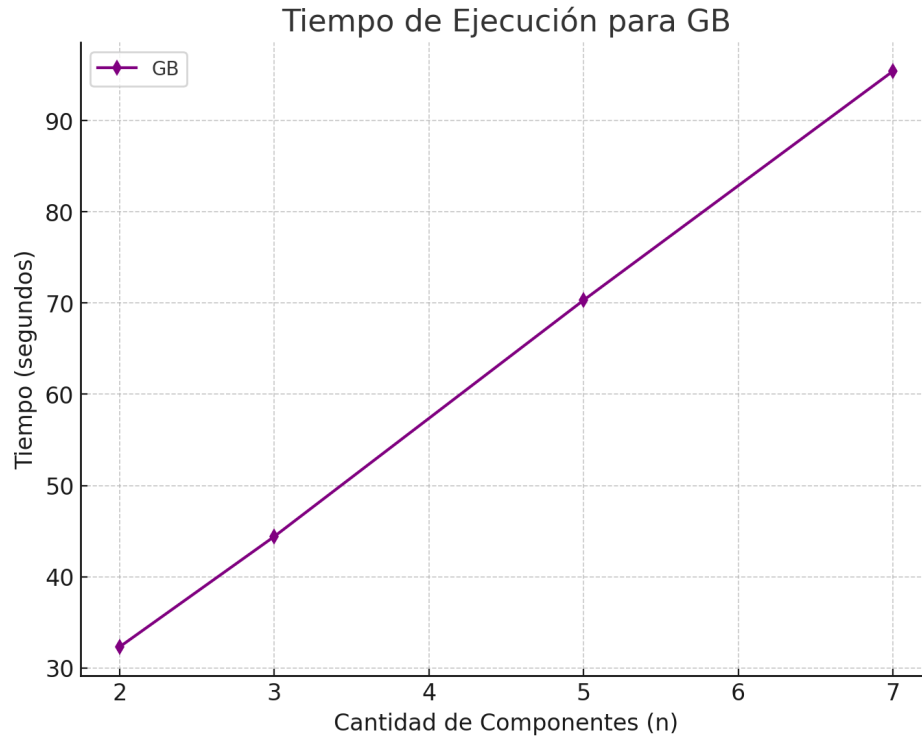


Figura 25: Grafico de Cantidad de Componentes vs tiempo que demoró entrenar el modelo
Fuente: Elaboración propia

Finalmente se pueden tomar los resultados de los análisis realizados para determinar que la combinación de PCA con XGB además de ser relativamente más rápida en comparación a las demás combinaciones, es la que presenta el menor **MSE**, por lo que se define que tomando $F = 1$, $C = 7$ junto a PCA + XGB se obtiene el mejor regresor posible.

4.3. Clasificador

Una vez teniendo la combinación que entrega los mejores resultados posibles, se comenzó el desarrollo de un pipeline que pueda realizar todo el proceso de preparación de los datos para poder luego alimentar un clasificador. A modo de resumen, los pasos de las secciones anteriores consisten en:

1. Cargar los videos transformados en JSON con los datos de expresiones faciales.
2. Dividir los conjuntos de datos por archivo en tercios y extraer el tercio del medio.
3. Reducir la lista de 57 expresiones utilizando PCA.
4. Dividir los datos en conjuntos de entrenamiento y prueba.

5. Entrenar XGBoost como regresor.
6. Exportar el regresor entrenado.

Un detalle importante es que una vez definidos los pasos que debe tomar el pipeline, se consideró dejar estáticos los componentes principales con $C = 7$ ya que esta fue la cantidad de componentes principales que entregó mejores resultados en las pruebas de C y F , sin embargo, no se descarta la idea de aumentar C en caso de ser necesario. Por otra parte, ya que $F = 1$ entregó el mejor resultado, no se puede disminuir más ya que se está tomando toda la información de los frames. Adicionalmente, debido a que predecir expresiones faciales es el foco central del proyecto subyacente al presente documento, el modelo fue probado exclusivamente con los gestos de **AbrirYCerrarBoca**, **ComisuraLabial**, **LevantarCejas** y **Pestañeo** (el cuál quedo con ese nombre debido a temas de encoding, pero claramente representa el pestañeo).

Para comenzar se recreó el comportamiento visto en la sección 4.1 en donde se tranforma la combinación de todos los archivos JSON almacenados en un dataframe, sin embargo, a diferencia del proceso descrito anteriormente, se generó una columna "label" donde se almacena el nombre del gesto/video del que es parte el frame. Posteriormente son filtrados todos aquellos datos que provienen de archivos que no sean expresiones faciales como tal, como lo son el movimiento lateral y vertical de la cabeza. Luego es generada una columna "label_encoded" la cual es una asignación de números a cada label en donde las asignaciones pueden ser vistas en la tabla 10.

Variable	Valor
AbrirYCerrarBoca	0
ComisuraLabial	1
LevantarCejas	2
Pestañeo	3

Tabla 10: Asignaciones de encoding para expresiones
Fuente: Elaboración propia

Para luego eliminar las columnas de **frame_index**, **file** y **label** dejando solamente **label_encoded**.

A continuación se crea una clase de python que pueda almacenar el regresor pre entrenado para poder emplearlo en el futuro, y como paso siguiente es creado un pipeline utilizando la función pipeline de Scikit Learn [Scikit-learn Developers, 2024b], la cual permite generar una secuencia de transformaciones de datos. El pipline emplea PCA con una cantidad variable de componentes principales para poder comparar el rendimiento, seguido por XGBoost pre entrenado y para finalizar se aplica SVC (Support Vector Classification), el cual es una implementación específica de SVM para un clasificador entregada por la biblioteca Scikit Learn [Scikit-learn Developers, 2024d]. El clasificador SVC necesita, al igual que los regresores vistos anteriormente, una separación de los conjuntos de entrenamiento y prueba, los cuales

son generados siempre con la misma semilla para poder obtener resultados reproducibles y en una razón de 80 - 20 respectivamente. Los datos además son separados manteniendo la proporción de clases en ambos conjuntos.

Más adelante, una vez que el pipeline está construido, fueron cargados los archivos de prueba, los cuales consisten en archivos JSON de videos los mismos cuatro movimientos mencionados anteriormente. Las columnas **label** y **label_encoded** también son agregadas a los datos de prueba para poder hacer una comparación del label asignado por el clasificador y el label verdadero del video. Finalmente, el modelo es evaluado utilizando el informe de evaluación entregado por Scikit Learn más una matriz de confusión para visualizar el rendimiento del clasificador.

Antes de continuar con el análisis y resultados de las pruebas realizadas es importante darle una breve definición a las métricas que serán usadas: precisión, recall y puntaje f1 (o f1-score en inglés). Estas tres métricas son entregadas en el informe de Scikit Learn.

La precisión se define como la relación entre los verdaderos positivos y todos los positivos entregados por el clasificador, o en palabras simples, entrega una medición de cuantas de las predicciones de una clase X son efectivamente clase X y no clase Y o Z . Por otra parte, el recall, o también llamado sensibilidad, entrega una relación entre los verdaderos positivos, y los elementos que pertenecen a la misma clase que los verdaderos positivos, o en otras palabras, entrega una medición para los elementos de la clase B que fueron catalogados correctamente y no como una clase distinta como A o C . El f1-score es el promedio armónico de las otras dos métricas mencionadas, por lo que entrega un resultado considerando tantos falsos positivos como falsos negativos, siendo útil cuando los datos de entrenamiento se encuentran sesgados. Sin embargo, como puede ser visto en la tabla 11, no existe un sesgo significativo en los datos por lo que el análisis se centrará en las primeras dos métricas mencionadas. Como una métrica adicional se considera el accuracy del modelo, que compara la cantidad de verdaderos positivos versus el total de datos, por lo que este será el indicador principal del rendimiento del clasificador.

Etiqueta	Frecuencia
3	376
2	351
1	326
0	320

Tabla 11: Distribución de etiquetas en los datos de entrenamiento
Fuente: Elaboración propia.

4.3.1. Análisis

Inicialmente fue considerado probar únicamente con $C = 7$ en vista de que fue la cantidad de componentes principales que entregó un mejor resultado con el regresor, sin embargo, se probó de igual manera con $C \in [2, 9]$, $C \in \mathcal{N}$, y no más de 9 porque, tomando en cuenta la idea detrás de la reducción de dimensionalidad, no tiene mucho sentido seguir aumentando la cantidad de componentes principales ya que mientras más componentes haya, menos información sobre el conjunto de datos entrega cada uno.

$C = 2$

Los resultados para $C = 2$ pueden ser observados en la tabla 12 y figura 26, además de tener una **accuracy de 44 %**. Esperablemente los resultados no son buenos ya que es difícil que solamente dos componentes puedan capturar la mayor parte de la información de 57 características, incluso a pesar de que hay algunas expresiones que no entregan mucha información como lo puede ser **eyeSquintLeft** (revisar tabla 16).

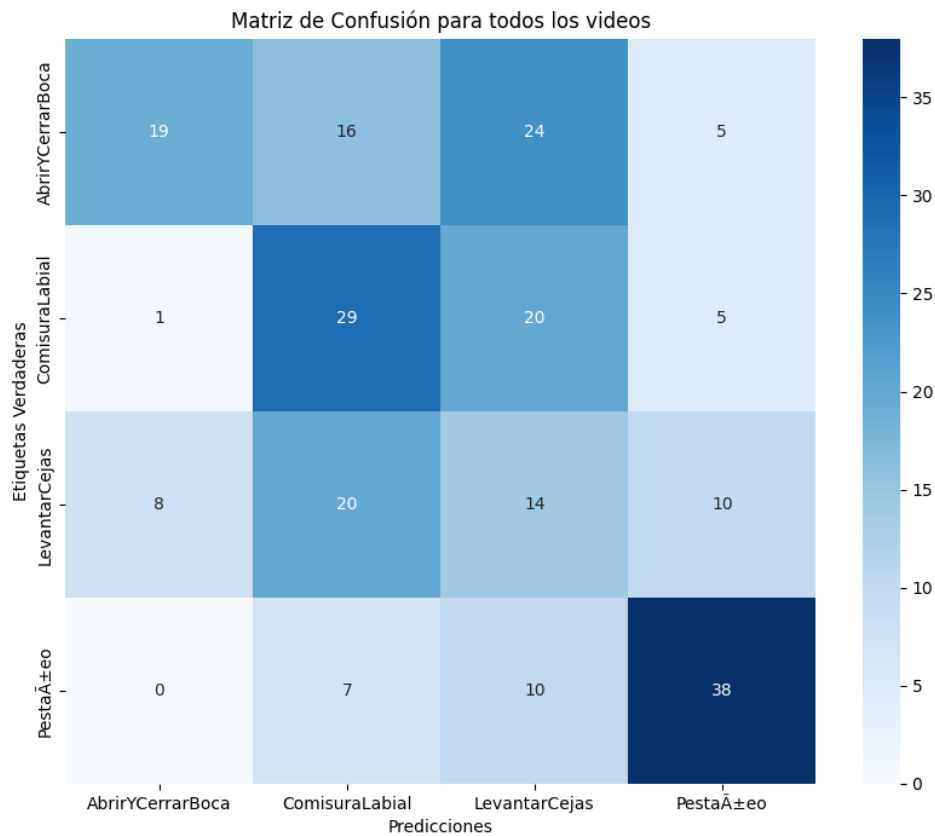


Figura 26: Matriz de confusión para $C = 2$
Fuente: Elaboración propia

Etiqueta	Precisión	Recall
AbrirYCerrarBoca	0.68	0.30
ComisuraLabial	0.40	0.53
LevantarCejas	0.21	0.27
Pestaño	0.66	0.69

Tabla 12: Resultados de precisión y recall por etiqueta para $C = 2$

$C = 3$

Los resultados para $C = 4$ pueden ser observados en la tabla 13 y figura 27, además de tener una **accuracy de 43%**. Este resultado es bastante interesante ya que, a pesar de que se puede apreciar una clara mejora en la clasificación de los gestos **AbrirYCerrarBoca**, **LevantarCejas** y **Pestaño**, el clasificador no acertó a ningún resultado del gesto **ComisuraLabial**, lo que resultó en que la accuracy disminuyera un 1%. Es importante mantener en mente este resultado ya que puede ser útil para una siguiente etapa del proyecto.

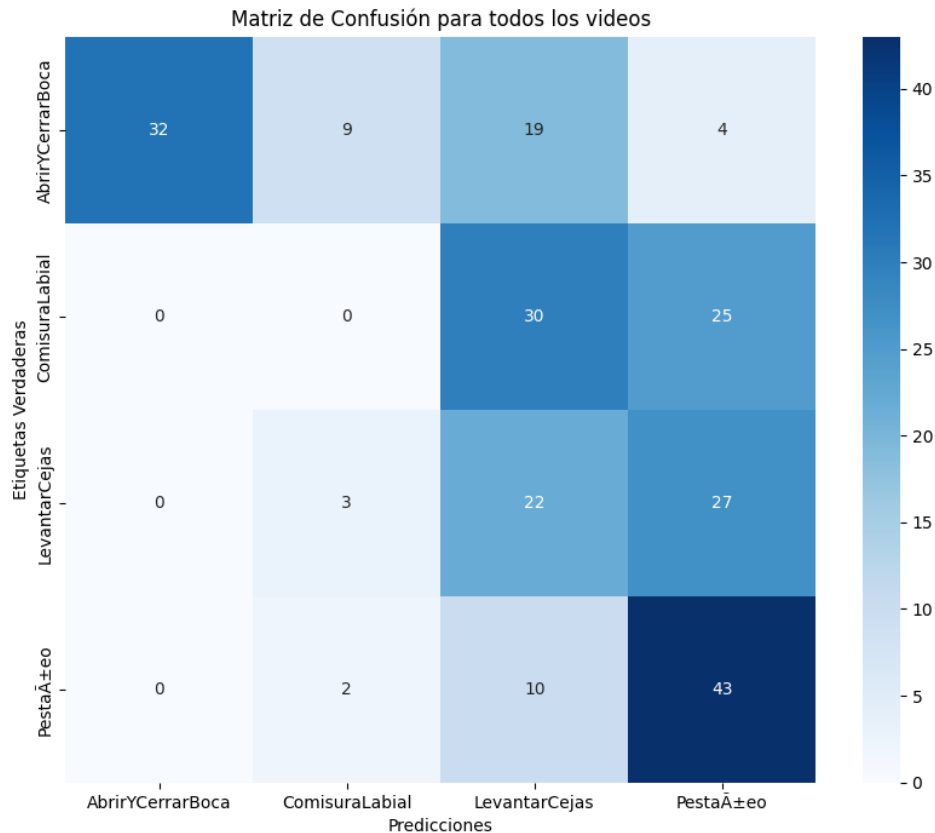


Figura 27: Matriz de confusión para $C = 3$

Fuente: Elaboración propia

Etiqueta	Precisión	Recall
AbrirYCerrarBoca	1.00	0.50
ComisuraLabial	0.00	0.00
LevantarCejas	0.27	0.42
Pestaño	0.43	0.78

Tabla 13: Resultados de precisión y recall por etiqueta para $C = 3$
Fuente: Elaboración propia

$C = 4$

Los resultados para $C = 4$ pueden ser observados en la tabla 13 y figura 27, además de tener una **accuracy de 59%**. En este paso se puede apreciar la primera mejora significativa en cuanto a clasificación aumentando casi en un 20%. Cabe mencionar que es sumamente llamativo que entre el paso anterior, en donde hubo 0 de precisión al clasificar comisura labial, aumentó a 0,54, sin embargo, el recall de **Pestaño** disminuyó. Hasta el momento $C = 4$ ha entregado el mejor resultado, por lo que el comportamiento se encuentra en línea con el descrito por el **MSE** del regresor.

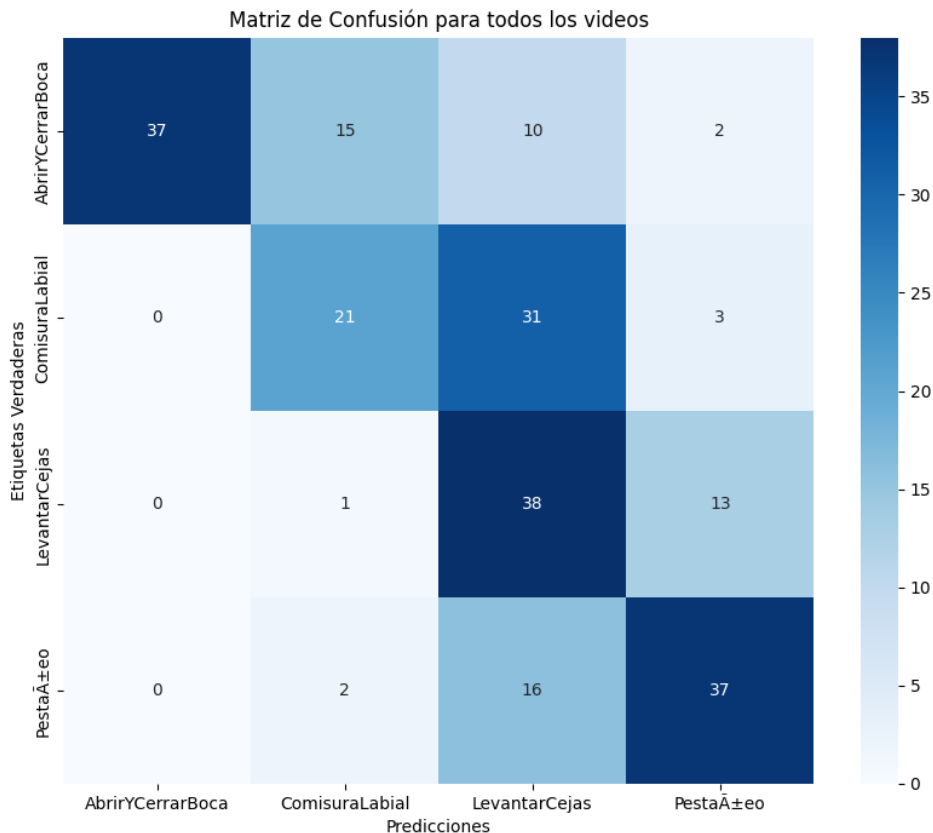


Figura 28: Matriz de confusión para $C = 4$
Fuente: Elaboración propia

Etiqueta	Precisión	Recall
AbrirYCerrarBoca	1.00	0.58
ComisuraLabial	0.54	0.38
LevantarCejas	0.40	0.73
Pestaño	0.67	0.67

Tabla 14: Resultados de precisión y recall por etiqueta para $C = 4$
Fuente: Elaboración propia

Entre $C = 5$ y $C = 8$, sin embargo, no hubo mejoras significativas como puede ser apreciado en el grafico 29, en donde el comportamiento observado en los pasos anteriores se repitió: alta precisión en **Pestaño** y **AbrirYCerrarBoca** mientras **ComisuraLabial** fue consistentemente el que tuvo peor desempeño.

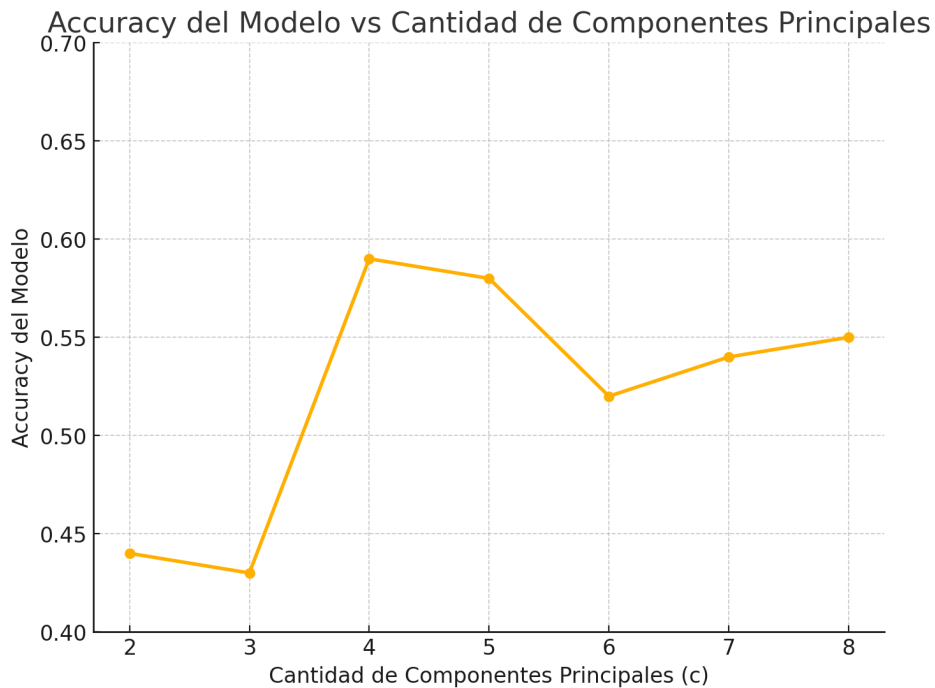


Figura 29: Accuracy vs cantidad de componentes principales
Fuente: Elaboración propia

Sin embargo, al usar $C = 9$ los resultados presentan una mejora por sobre cualquier otro C . Como puede ser visto en la figura 30 y tabla 15. Con un **accuracy de 62%** es el que presenta los mejores resultados, incluso teniendo una buena precisión en **ComisuraLabial**.

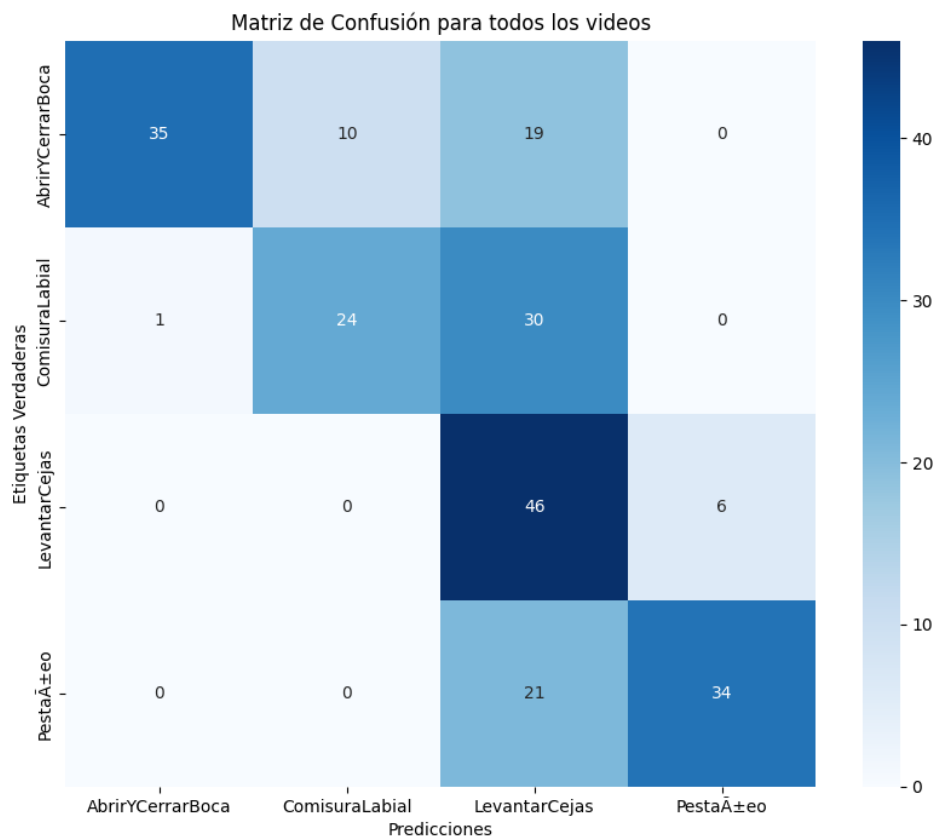


Figura 30: Matriz de confusión para $C = 9$
Fuente: Elaboración propia

Etiqueta	Precisión	Recall
AbrirYCerrarBoca	0.97	0.55
ComisuraLabial	0.71	0.44
LevantarCejas	0.40	0.88
Pestaño	0.85	0.62

Tabla 15: Resultados de precisión y recall por etiqueta

CAPÍTULO 5

CONCLUSIONES

El presente documento ha abordado la implementación de tecnologías para la creación de un clasificador de gestos faciales con el fin de poder implementarlo en una situación de hospitalización clínica o domiciliaria para pacientes con discapacidad motora severa. Durante las etapas del desarrollo de la plataforma y el modelo de clasificación se consideró de suma importancia la escalabilidad del mismo con el fin de que el modelo pueda seguir mejorando a lo largo del tiempo al añadir una mayor cantidad de gestos, al mismo tiempo de recopilar más datos para el pre-entrenamiento del modelo.

Del comportamiento del clasificador se pueden entender ciertos problemas que pudieron afectar negativamente el resultado, por ejemplo, para la gran mayoría de pruebas que fueron realizadas, hubo una fuerte relación entre el levantamiento de cejas y el pestañeo, lo que se puede deber a que en general las personas tienden a pestañear mientras levantan las cejas. Por otra parte, fue observado durante la toma de videos (como opinión personal del autor) que el movimiento que más complicó a los voluntarios fue el de las comisuras de la boca ya que muchos no podían hacer un movimiento muy pronunciado. Sin embargo, esta dificultad para detectar ciertos gestos puede remediarse con la posesión un conjunto de datos más extenso, por lo que se plantea seguir utilizando el sistema de recolección de datos para, en un trabajo futuro por parte del equipo del proyecto, realizar el mismo análisis extensivo y comparar así los resultados.

Por otra parte, el comportamiento que muestra el clasificador también sirve para tener en consideración los gestos que serán detectados por el clasificador en la implementación clínica del prototipo del proyecto. Se aconseja, por ejemplo, que el movimiento de apertura de boca se relacione con algún movimiento que pueda llegar a ser repetitivo, ya que además de ser sencillo de realizar, es un movimiento primordial para la vida (gritar, comer, hablar, bostezar, etc).

Además de lo mencionado anteriormente se presenta el siguiente trabajo a futuro por realizar en siguientes iteraciones del proyecto.

Trabajo a Futuro

El trabajo realizado para el presente documento se realizó pensando siempre en posibles mejoras futuras al modelo, las cuales por limitaciones de tiempo o complejidad no pudieron ser llevadas a cabo en esta etapa del proyecto.

Durante el desarrollo del portal de captura de videos se ideó la primera mejora, la cual causó que la plataforma de toma de videos fuera diseñada con el objetivo de agregar

nuevas capturas de gestos de manera sencilla. Esta mejora apunta a la ampliación de movimientos o gestos a capturar en el futuro tales como lo pueden ser el cerrar un solo ojo, movimientos de boca como un 'pucker' o "boca de pato", o incluso gestos más complejos como arrugar la cara (similar a la expresión de una persona luego de comer un limón). Además, debido a que los label que ocupa el clasificador son creados en base al mismo video, la adición de nuevos gestos no afectará al clasificador

La segunda mejora se presentó durante el proceso de extracción de datos de los videos, en donde para cada frame son almacenados archivos JSON con la información pertinente a la probabilidad de expresiones dictada por Mediapipe, junto a las coordenadas de todos los puntos del facemesh que entrega el mismo. Para esta iteración del clasificador se está tomando en cuenta solamente las expresiones, por lo que se podría integrar la información de los más de 400 puntos entregados por el facemesh para que el clasificador pueda trabajar con una mayor cantidad de información al enfrentar el problema.

En la misma línea de la mejora anterior, se propone para una siguiente iteración la identificación de componentes principales para cada uno de los gestos de manera individual, con el fin de tener un mayor grado de precisión en el clasificador, sin embargo esto requiere de un mayor nivel de datos por lo que será implementado en la medida que estos sean capturados.

Se propone, además, el aumento de datos para el entrenamiento del regresor ya que así se podría probar con una granularidad de frames mayor ya que el **MSE** del regresor presentó una relación de proporcionalidad con la cantidad de frames utilizados para entrenar al mismo.

ANEXOS

Vistas de aplicación web

Las vistas del formulario son las siguientes

Terms of Service

Enter your details:

Nombre:

Apellido:

Edad:

Sexo:

Enfermedad de base:

Peso estimado (kg):

Altura (cm):

Figura 31: : Vista de Formulario.
Fuente: Elaboración propia.

MODELO ANTERIOR

El modelo anterior fue creado utilizando Mediapipe de Google. Mediapipe es una colección de bibliotecas y herramientas enfocadas en la implementación rápida de inteligencia artificial y aprendizaje de máquina en desarrollos de aplicaciones. El conjunto de soluciones brindadas por Mediapipe cubre un amplio rango desde generación de imágenes y clasificación de texto, hasta reconocimiento de gestos y clasificación de audio [AI, 2024]. En este caso, MediaPipe fue utilizado para:

- Detección facial: a pesar de que MediaPipe puede detectar caras y gestos en más de una cara al mismo tiempo en una imagen, debido a que el objetivo del proyecto es brindarle asistencia a los pacientes en donde cada uno tenga un "control" separado, se decidió solamente procesar una cara a la vez, por lo que si aparece una segunda cara en el feed (ya sea de video o de cámara) no se toma en cuenta.

- Extracción de Face Mesh⁵ y de Gestos: Luego de identificar la cara objetivo, se utilizó el Face Mesh para poder extraer la posición de todos los puntos en la cara y l

Posteriormente, se formó una secuencia con los datos de cada frame y en base a metaheurísticas definidas por el desarrollador del modelo se definieron conjuntos de secuencias que son detectadas como gestos.

Consentimiento informado

⁵El Face Mesh de MediaPipe es una solución que estima 468 puntos de referencia faciales en tiempo real.[MediaPipe, 2024]

CONSENTIMIENTO INFORMADO

“Interfaz humano-máquina para la asistencia y cuidado de personas mayores con dependencia motora: una propuesta transformadora basada en visión computacional y videoanalítica”

Proyecto financiado por ANID

DESAFÍO GLOBAL DE LONGEVIDAD SALUDABLE AÑO 2023 FASE INTERNACIONAL CATALIZADORA - INVESTIGACIÓN APLICADA

Folio: NAM23I0021

Nombre del Investigador principal: Claudio Tapia Malebrán

RUT.: 12440826-1

Institución: Universidad de Chile.

Dirección: Avenida Independencia 1027

Correo Electrónico: claudiotapia@uchile.cl

Teléfono: +56984099523

Invitación a participar: Le estamos invitando a participar en el proyecto de investigación “Interfaz humano-máquina para la asistencia y cuidado de personas mayores con dependencia motora: una propuesta transformadora basada en visión computacional y videoanalítica” debido a que usted cumple con los criterios para participar en este proyecto de investigación.

Objetivos: Este proyecto tiene como objetivo desarrollar un sistema de comunicación entre una persona y diferentes dispositivos electrónicos (como encender y apagar la luz, solicitar ayuda a través de un timbre, encender y apagar el televisor, entre otros), a partir de las imágenes de video registradas por una cámara. Esto permitirá, en el futuro, la detección de gestos de su rostro o lenguaje corporal, para la asistencia, cuidado y monitoreo de personas mayores con dependencia motora. Este proyecto tiene el potencial de aumentar la eficiencia de los sistemas de cuidado al mejorar la autonomía de las personas mayores con discapacidades motoras, disminuir los efectos negativos de la dependencia, como la depresión o el deterioro de la salud mental, y mejorar la calidad de vida de las personas mayores.

¿En qué consiste su participación?

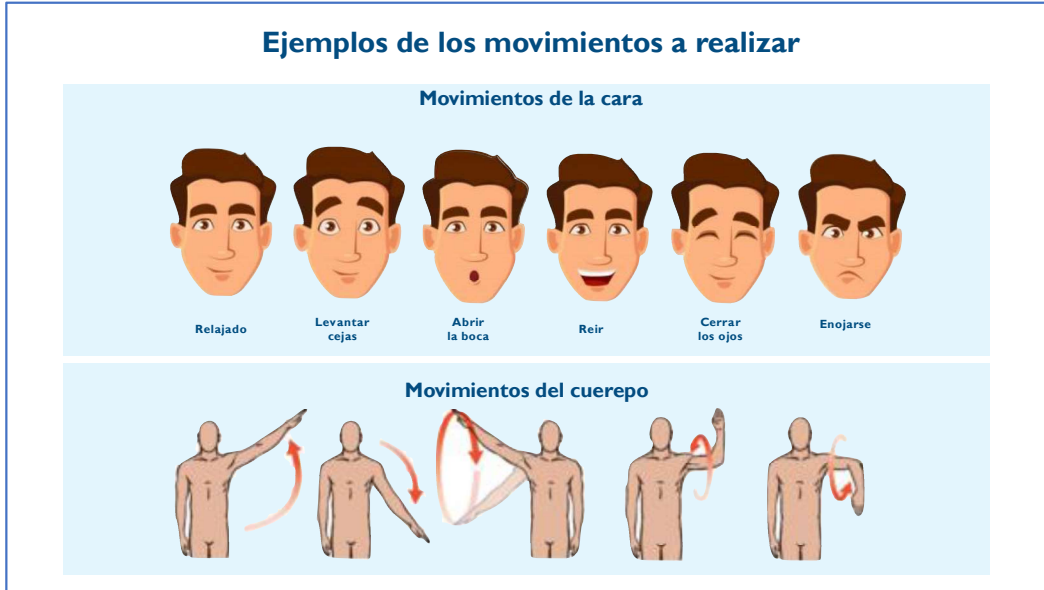
Su participación implica los siguientes procedimientos:

1. Se consultarán sus antecedentes generales de salud. También el equipo de investigación llevará a cabo una evaluación de su nivel de movilidad utilizando un test llamado Escala de Evaluación Motora. Esta escala incluirá la realización de movimientos cotidianos como girarse en la cama, sentarse al borde de la cama, mantenerse de pie, caminar y manipular objetos. Además se realizará una evaluación conocida como Mini-Mental que nos ayudará a saber cómo está su memoria y función cognitiva.

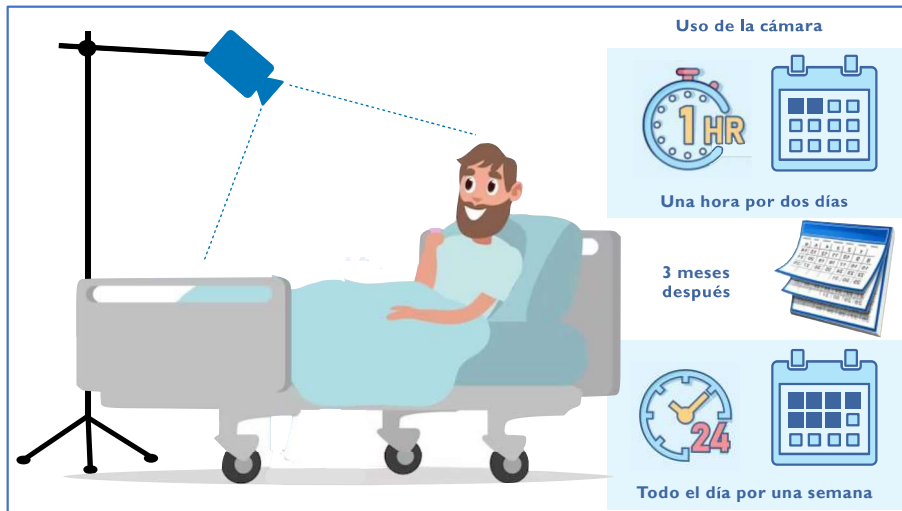
12|OCT|2023



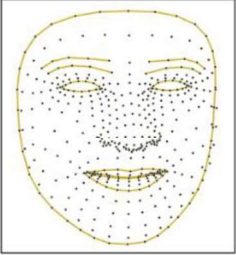

2. Se posicionará frente a su cama una cámara de video de tamaño pequeño que registrará algunos gestos de su cara y/o movimientos corporales que usted pueda realizar. Esto no tiene costo para usted.



Su participación se realizará en dos oportunidades. La primera vez (fase 1), se realizará una sesión de una hora durante dos días. La segunda (fase 2), se realizará 3 meses después y durará una semana, en ese periodo la cámara estará funcionando todo el día. No se almacenará ninguna imagen o video de su rostro, de su cuerpo o de su habitación.



A continuación se muestra en qué consistirá su participación:

Participación	¿Qué ocurrirá?	Requisitos	¿Me grabarán?	¿Dónde se guardarán los datos?
<p>Fase 1</p> <p>Día 1 y 2 Duración: 1 hora cada día.</p>	<p>Uno de los investigadores instalará una cámara pequeña en su habitación y le pedirá que haga movimientos con la cara y el cuerpo (como los de la figura más arriba) para que la computadora pueda identificar estos gestos. Al final del segundo día se retirará la cámara.</p>	<p>Debe estar acompañado/a por un cuidador/a en todo momento.</p>	<p>La cámara no guardará ninguna grabación de imágenes, sólo permitirá identificar una especie de esquema sobre sus gestos y movimientos, como se ve en la figura.</p> 	<p>Serán almacenados de manera segura en una cuenta privada en internet propiedad del investigador y administrada por la Universidad de Chile.</p>
<p>Fase 2</p> <p>7 días Duración: 1 semana (tres meses después de la primera participación)</p>	<p>Uno de los investigadores instalará una cámara pequeña en su habitación. Usted no requiere realizar ninguna acción y podrá seguir con todas sus actividades habituales. La cámara de video estará permanentemente registrando sus gestos faciales y/o movimientos corporales para detectar su necesidad de asistencia. Usted tiene la posibilidad de apagar la cámara o interrumpir el monitoreo en cualquier momento.</p>	<p>Debe estar acompañado/a por un cuidador/a en todo momento.</p>		

- Al finalizar su participación y luego de que la cámara de video sea retirada, un investigador le realizará algunas preguntas acerca de su percepción del procedimiento en general y de su percepción acerca del uso de las prestaciones realizadas a partir del uso de la cámara. El investigador registrará sus respuestas. Usted podrá revisar estas respuestas en ese mismo momento y también podrá solicitar una copia de este documento.

Riesgos: Debido al tipo de estudio y basado en la evidencia disponible hasta la fecha, no existe información que nos haga prever algún tipo de riesgo al participar de este estudio. De todas formas, el equipo investigador estará permanentemente consultando por su estado, y ante cualquier sospecha de presentar un riesgo asociado a su participación en el proyecto, el equipo de investigadores suspenderá de manera segura los procedimientos y su participación.

Sólo se podrán realizar las etapas del estudio si usted se encuentra acompañado/a o bajo los cuidados de un cuidador o cuidadora. Es importante destacar que el sistema no implica contacto directo con su piel o resto del cuerpo.

Costos: Los implementos necesarios para la realización de esta investigación serán aportados por el equipo de investigación, sin costo alguno para usted durante el desarrollo de este proceso investigativo.

Beneficios: Su participación en este estudio contribuirá a desarrollar un sistema de alerta y monitoreo que pueda ser utilizado por personas con movilidad reducida. Al mismo tiempo, su participación contribuirá a identificar un conjunto de variables que pueden ser usadas en dispositivos tecnológicos que detecten automáticamente la necesidad de asistencia por parte de un tercero. El contar con este tipo de información permitiría mejorar la toma de decisiones por parte del equipo de salud.

Compensación: Usted no recibirá ninguna compensación económica por su participación en el estudio. Tampoco recibirá algún beneficio directo de este estudio.

Confidencialidad: Los resultados que se obtengan de este estudio servirán para realizar un reporte científico que describa el impacto de contar con un sistema automático de asistencia para personas con movilidad reducida. En estos reportes la identidad de los participantes no será publicada, resguardando así su privacidad. La integridad de los datos obtenidos será almacenada en formato digital (servicio de almacenamiento con cuentas de la Universidad de Chile OneDrive) mientras dure el estudio y posteriormente los datos se mantendrán al resguardo del investigador responsable en la Universidad de Chile.

Es posible que los datos recopilados en el marco de esta investigación sean utilizados en estudios posteriores que se beneficien del tipo de registros obtenidos. Si así fuese, se almacenarán por un periodo de 10 años y solamente estarán disponibles los datos codificados según lo indicado en el párrafo anterior, manteniendo su identidad personal estrictamente en el anonimato.

Usted puede decidir si está de acuerdo en autorizar que sus datos sean usados para futuros estudios.

Usted puede solicitar al investigador responsable que destruya los datos provenientes de su participación una vez finalizado este proyecto, si así usted lo desea.

Información adicional: Usted y el equipo de salud tratante serán informados si surgen complicaciones que puedan afectar su participación en el proceso investigativo.

No se espera que los participantes se queden con el dispositivo utilizado en este estudio, ni que haya beneficios directos para ellos. Los resultados de la investigación serán utilizados únicamente para fines científicos.

Voluntariedad: Su participación en esta investigación es totalmente voluntaria y se puede retirar en cualquier momento comunicándole al investigador, sin que ello signifique modificaciones en el estudio y tratamiento habituales de su condición de salud. De igual manera el investigador podrá determinar su retiro del estudio si considera que esa decisión va en su beneficio.

Complicaciones: En el improbable caso de que Ud. presente alguna complicación directamente dependientes del uso del sistema de la cámara de video, su participación será suspendida y usted recibirá una evaluación médica o psicológica por parte de un profesional del Servicio de Salud Metropolitano Oriente, sin costo alguno para usted o su sistema previsional. Esto no incluye las complicaciones propias de su enfermedad y de su curso natural.

12|OCT|2023



Derechos del Participante: Usted tiene derecho a ser informado de todo lo que se realizará en todo momento, pudiendo hacer las preguntas que sean necesarias para la mejor comprensión de los procedimientos.

Usted tiene el derecho a solicitar toda la información recopilada a partir de su participación. En caso de requerirlo, será proporcionada una vez entregado el informe final del proyecto, esto es 12 meses después de iniciado. La información será entregada en formato digital a través de correo electrónico conteniendo un enlace (link) que permitirá descargar toda los datos e información asociada a su participación, o bien será entregada en formato físico si así usted lo desea.

Usted recibirá una copia íntegra y escrita de este documento firmado. Si usted requiere cualquier otra información sobre su participación en este estudio puede comunicarse con:

Investigador: Claudio Tapia Malebrán / +56 9 84099523

Otros Derechos del participante

En caso de duda sobre sus derechos debe comunicarse con el Presidente del “Comité de Ética de Investigación en Seres Humanos”, Dra. Lucía Cifuentes, al Email: ceish.med@uchile.cl. cuya oficina se encuentra en la Facultad de Medicina de la Universidad de Chile, Avenida Independencia 1027, teléfono: (+56 2) 2 9789536.

Conclusión:

Después de haber recibido y comprendido la información de este documento y de haber podido aclarar todas mis dudas, otorgo mi consentimiento para participar en la investigación “Interfaz humano-máquina para la asistencia y cuidado de personas mayores con dependencia motora: una propuesta transformadora basada en visión computacional y videoanalítica”.

¿Está de acuerdo en autorizar que sus registros sean usados para futuros estudios?

Si su respuesta es “Si”, sólo se usarán los datos relacionados con las ubicaciones específicas de puntos clave en el rostro y la postura corporal, omitiendo otros datos. Estos datos serán usados para estudios de optimización y mejora de los sistemas de detección movimientos del rostro o del cuerpo, para la asistencia, cuidado y monitoreo de personas mayores.

En ningún caso serán usadas imágenes, pues ninguna imagen será almacenada en este estudio.

Si _____

No _____

12|OCT|2023



Nombre del Participante
Rut.

Firma

Fecha

Nombre de Director
de Institución o Delegado
Art. 11 Ley 20120
Rut.

Firma

Fecha

Nombre del investigador
Rut.

Firma

Fecha

12|OCT|2023



Selecciona una acción para grabar

Al seleccionar una de las acciones serás llevado a una página en donde se tomarán los videos. Una vez dentro puedes ver el video de ejemplo para saber que hacer, luego podrás grabar el video. Durante la grabación del video no podrás ver la cámara pero si después. Para iniciar y detener la grabación debes presionar el botón correspondiente. Una vez terminada la grabación podrás ver el video recién tomado, si no te gusta podrás tomar otro, y si estás conforme puedes volver al menú para el siguiente gesto. Una vez que todos los gestos sean grabados, puedes presionar el botón de finalizar.

Movimiento Arriba/Abajo
Movimiento Lateral
Levantar Cejas
Abrir/Cerrar Boca
Movimientos de Comisura Labial
Pestañeo
Posición de la Cabeza

Finalizar

Figura 32: : Vista de menú.
Fuente: Elaboración propia.

¡Gracias por tu colaboración!
Muchas gracias por su colaboración, ya puede cerrar esta pestaña.

Figura 33: : Mensaje de salida.
Fuente: Elaboración propia.

Lista de expresiones

Expresión	Expresión	Expresión
_neutral	cheekPuff	mouthClose
browDownLeft	cheekSquintLeft	mouthDimpleLeft
browDownRight	cheekSquintRight	mouthDimpleRight
browInnerUp	eyeBlinkLeft	mouthFrownLeft
browOuterUpLeft	eyeBlinkRight	mouthFrownRight
browOuterUpRight	eyeLookDownLeft	mouthFunnel
eyeLookInLeft	eyeLookDownRight	mouthLeft
eyeLookInRight	eyeLookUpLeft	mouthLowerDownLeft
eyeLookOutLeft	eyeLookUpRight	mouthLowerDownRight
eyeLookOutRight	eyeSquintLeft	mouthPressLeft
eyeSquintRight	eyeSquintRight	mouthPressRight
eyeWideLeft	eyeWideRight	mouthPucker
eyeWideRight	jawForward	mouthRight
jawForward	jawLeft	mouthRollLower
jawLeft	jawOpen	mouthRollUpper
jawOpen	jawRight	mouthShrugLower
jawRight	mouthShrugUpper	mouthUpperUpLeft
mouthSmileLeft	mouthSmileRight	mouthStretchLeft
mouthStretchRight	noseSneerLeft	noseSneerRight

Tabla 16: Lista de Expresiones

REFERENCIAS BIBLIOGRÁFICAS

[AI, 2024] AI, G. (2024). Mediapipe solutions guide.

[AWS, 2024a] AWS (2024a). Amazon ec2 instance types.

[AWS, 2024b] AWS (2024b). Amazon ec2 instance types - t2.

[AWS, 2024c] AWS (2024c). Amazon route 53.

[Brenning, 2023] Brenning, A. (2023). Interpreting machine-learning models in transformed feature space with an application to remote-sensing classification. *Machine Learning*, 112(9):3455–3471.

[García et al., 2016] García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M., y Herrera, F. (2016). Big data preprocessing: methods and prospects. *Big Data Analytics*, 1(1):1–22.

[MDN, 2024] MDN (2024). Web manifest.

[MediaPipe, 2024] MediaPipe (2024). Face mesh.

[NGINX, 2024] NGINX, I. (2024). Nginx documentation.

[npm, 2024a] npm, I. (2024a). package.json.

[npm, 2024b] npm, I. (2024b). What is npm?

[React.js, 2024] React.js (2024). Create react app documentation.

[Scikit-learn Developers, 2024a] Scikit-learn Developers (2024a). sklearn.model_selection.train_test_split. https://scikit-learn.org/1.5/modules/generated/sklearn.model_selection.train_test_split.html.

[Scikit-learn Developers, 2024b] Scikit-learn Developers (2024b). sklearn.pipeline.pipeline. <https://scikit-learn.org/1.5/modules/generated/sklearn.pipeline.Pipeline.html>.

[Scikit-learn Developers, 2024c] Scikit-learn Developers (2024c). sklearn.svm.svc. <https://scikit-learn.org/1.5/modules/generated/sklearn.svm.SVC.html>.

[Services, 2024] Services, A. W. (2024). Amazon simple storage service (s3).

[Wikipedia, 2024a] Wikipedia (2024a). Favicon.

[Wikipedia, 2024b] Wikipedia (2024b). Tienda de software.

[World Health Organization, 2021] World Health Organization (2021). *Global strategy on digital health 2020-2025*. World Health Organization, Geneva. Licence: CC BY-NC-SA 3.0 IGO.