

## PROGRAMA DE INCENTIVOS PARA LA FORMACIÓN DE DOCENTES-INVESTIGADORES

Nombre del programa de posgrado: Maestría en Ciencias de la Computación Categorización PRONII: --

Nombre de la Institución: Facultad Politécnica - UNA

Vinculación a Proyectos I+D: --

Nombre del beneficiario: Carolina Elizabeth Villegas Colmán

Vinculación docencia, tutoría o centro de investigación: Núcleo de Investigación y Desarrollo Tecnológico

Publicaciones realizadas durante el programa: Conferencia Ibero-Americana de Computación Aplicada 2021

Título de tesis: Análisis Exploratorio de las Relaciones Lineales y no Lineales de las señales de Fotopleletismografía Aplicando Análisis de Componentes Principales para la Estimación de la Presión Arterial.

### RESUMEN

Las señales de fotopleletismografía en modelos de aprendizaje automático para la estimación de la presión arterial pueden verse afectadas cuando el conjunto de atributos es de alta dimensión. Reducir la dimensión de los atributos, eliminando aquellos atributos redundantes podrían ayudar a mejorar el rendimiento de los modelos predictivos. Este trabajo propone explorar las relaciones multivariadas lineales y no lineales de las señales de fotopleletismografía. Esto se consigue aplicando análisis de componentes principales mediante la descomposición de las matrices de correlaciones de Pearson y Spearman como técnica de reducción de dimensionalidad. Los resultados muestran que la matriz de correlación de Spearman obtiene levemente una mayor varianza acumulada en los primeros componentes principales. Las predicciones de la presión arterial con los nuevos componentes principales generados para los modelos predictivos satisfacen los estándares de la Asociación para el Avance de la Instrumentación Médica.

#### OBJETIVOS

##### Objetivo General:

Reducir la dimensionalidad del conjunto de datos explorando las relaciones multivariadas lineales y no lineales de las señales de fotopleletismografía a través del análisis de componentes principales.

##### Objetivos Específicos:

- Explorar las relaciones multivariadas lineales y no lineales de las señales de fotopleletismografía aplicando las correlaciones de Pearson y Spearman.
- Determinar las varianzas explicadas y acumuladas de los componentes principales aplicando análisis de componentes principales mediante la descomposición de las matrices de correlaciones de Pearson y Spearman.
- Seleccionar los componentes principales que retengan porcentajes de varianzas acumuladas con una mínima pérdida de información.
- Evaluar el rendimiento de los modelos predictivos entrenados con los componentes principales seleccionados para la estimación de la presión arterial.

#### APORTES DE LA INVESTIGACIÓN

Las contribuciones son:

- Reducir la dimensionalidad del conjunto de datos seleccionando los componentes principales de las señales de fotopleletismografía con la aplicación del análisis de componentes principales mediante la descomposición de las matrices de correlaciones de Pearson y Spearman.

- Implementar los nuevos componentes principales para mejorar el rendimiento de los modelos predictivos en la estimación de la presión arterial.

#### ACTIVIDADES REALIZADAS

A continuación se presenta el proceso de estimación de la presión arterial.

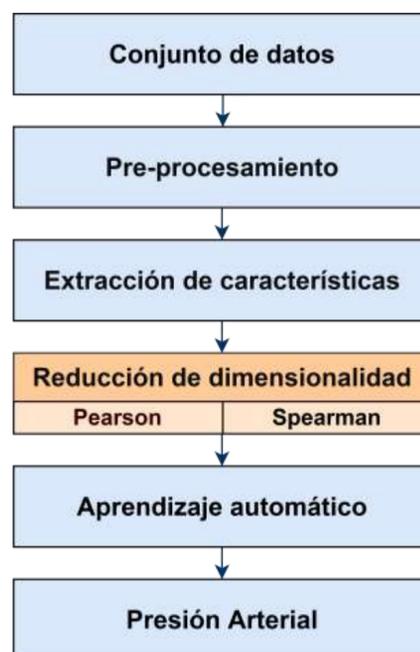


Figura 1. Proceso de estimación de la presión arterial.

#### RESULTADOS OBTENIDOS

Se realizaron 3 experimentos. El primero correspondió a la exploración de las relaciones multivariadas lineales y no lineales. En el segundo, se determinaron los porcentajes de varianzas acumuladas de los componentes principales y se redujeron las dimensiones del conjunto de datos.

Y finalmente, se analizaron los resultados obtenidos de los modelos predictivos con los componentes principales seleccionados.

#### CONCLUSIÓN

Se redujo el conjunto de datos de alta dimensión mediante la aplicación del análisis de componentes principales teniendo en cuenta las correlaciones de Pearson y Spearman. La matriz de correlación de Spearman presenta levemente una mayor varianza acumulada desde el primer componente principal con respecto al de Pearson. Para el entrenamiento de los modelos predictivos se seleccionaron 2, 3, 5 y 8 componentes principales, los cuales retenían el 70%, 80%, 90% y 96% de varianzas acumuladas con respecto al total. Los modelos predictivos con mejores rendimientos fueron los que se entrenaron con los componentes principales reducidos a través del análisis de componentes principales mediante la descomposición de la matriz de correlación de Spearman. Los modelos predictivos entrenados con 5 componentes principales obtuvieron mejores rendimientos.

#### VISIÓN Y PLANES FUTUROS

Para trabajos futuros se recomienda seleccionar diferentes tipos de señales de fotopleletismografía, y realizar un análisis comparativo con otras técnicas de reducción de atributos.

“Este programa de posgrado fue cofinanciado por el Consejo Nacional de Ciencia y Tecnología - CONACYT con recursos del FEEI”