**Abstract— This article exposes the procedure to develop a machine learning application using cloud computing tools. It then shows how to implement the model as a web service that can be accessed by other services using a microservices architecture. In particular, the application that is presented has the objective of predicting heart failure, taking into account the importance of health care today.**

**Keywords: Machine Learning, Cloud Computing, Heart Failure, Prediction, Web Services.**

1. INTRODUCCIÓN

La adaptación de la computación en la nube por parte de pequeñas y grandes empresas alrededor del mundo, ha traído una nueva generación de soluciones y propuestas tecnológicas por parte de compañías que han logrado reducir costos haciendo uso de arquitecturas en la nube, escalar productos de manera más efectiva, garantizar la disponibilidad de los servicios e innovar con productos de TI que generan valor a los usuarios. [1]

Entre las posibilidades que ofrece la computación en la nube se identifica la facilidad de incorporar tecnologías emergentes de una manera sencilla y costo efectiva, la incorporación de tecnologías como el Machine Learning (ML) para el desarrollo de servicios en distintos ámbitos ha sido posible gracias al desarrollo de la computación en la nube, actualmente es muy posible producir de manera efectiva soluciones que incorporen esta y otras tecnologías.

La medicina y salud en general han sido algunos de los contextos en los que se ha posibilitado el desarrollo de soluciones tecnológicas haciendo uso de diversas tecnologías como el ML, la capacidad para procesar datos y obtener una respuesta de un modelo de clasificación, regresión u otro, puede permitir la detección temprana de patologías y la confirmación de dictámenes médicos de forma rápida. [2]

1. MARCO TEÓRICO
   1. MACHINE LEARNING

El Machine Learning o Aprendizaje Automático es un enfoque de la inteligencia artificial que permite desarrollar algoritmos para predecir eventos u obtener respuestas a decisiones complejas a partir del procesamiento de datos y la identificación de patrones. Para lograr estos resultados es necesario que las máquinas pasen por un proceso de aprendizaje en el que iteran múltiples veces sobre datos de entrenamiento y aprenden de los errores obtenidos, existen múltiples enfoques de ML entre los que se encuentran el aprendizaje supervisado, no supervisado, semi-supervisado y por refuerzo. [3]

Para entender el proceso que sigue un algoritmo de ML es necesario entender alguna terminología relacionada al concepto.

*Dataset:* Un conjunto de datos que contiene ejemplos de cómo resolver el problema deseado.

*Características:* Datos clave que ayudan a entender un problema, son usados por los algoritmos de ML para aprender.

*Modelo de ML:* Es el resultado obtenido posterior al aprendizaje de un algoritmo de ML.

El proceso que sigue un algoritmo de ML para aprender y desplegarse es, la recopilación de datos, el preprocesamiento de los datos, el entrenamiento del modelo a partir de los datos entregados, la evaluación del modelo obtenido y el ajuste del modelo para optimizar el rendimiento.

* 1. COMPUTACIÓN EN LA NUBE

La computación en la nube es un modelo para facilitar el acceso a recursos computacionales como redes, servidores, almacenamiento y servicios, bajo demanda, algunas ventajas que ofrece este modelo son la capacidad para escalar rápidamente soluciones de TI, eficiencia operacional al tercerizar la administración de los recursos de TI y la reducción de costos en aspectos de TI al pagar sólo por los recursos usados bajo demanda. [4]

La computación en la nube ha tenido un gran impacto en la industria brindando a las pequeñas y grandes empresas la posibilidad de acelerar los tiempos de despliegue de soluciones, acelerar el crecimiento de las compañías y reducir costos en adquisición y mantenimiento de TI.

Algunos modelos de despliegue que se logran identificar en la computación en la nube son:

*Nube privada:* Infraestructura que se implementa para el uso exclusivo de una organización.

*Nube comunitaria:* Infraestructura que se implementa para el uso de un conjunto de organizaciones o consumidores con necesidades y características similares.

*Nube pública:* Infraestructura que se implementa para el uso abierto del público en general.

*Nube híbrida:* Infraestructura que se implementa y combina la nube privada y pública.

Los modelos de servicio a través de los cuales se puede acceder a servicios de computación en la nube son: Software como servicio (SaaS) que entrega una solución de software en la nube completa y lista para ser consumida, Plataforma como servicio (PaaS) que brinda una plataforma para desarrollar una solución software personalizada e Infraestructura como servicio (IaaS) que entrega un mecanismo para ejecutar servidores virtuales en la nube.

* 1. ENFERMEDADES CARDIOVASCULARES

Las enfermedades cardiovasculares (ECV) son un grupo de desórdenes del corazón y de los vasos sanguíneos que representan la principal causa de muerte a nivel mundial. De acuerdo con la Organización Mundial de la Salud (OMS) cada año se registran más muertes a causa de ECV que de otras enfermedades [5].

Adicionalmente, se tiene la creencia que la mayoría de estas enfermedades cardiovasculares pueden prevenirse si se actúa sobre factores de riesgo asociados al comportamiento de las personas, como el tipo de alimentación, actividad física, consumo de alcohol y tabaco [5].

1. DESCRIPCIÓN DEL PROBLEMA

Las enfermedades cardiovasculares (ECV) son la principal causa de muerte a nivel mundial, se estima que anualmente cerca de 17.9 millones de personas fallecen debido a ECV en todo el mundo, lo que representa el 31% de las muertes a nivel mundial. El 85% de estas muertes se deben a infartos y accidentes cerebrovasculares, y aproximadamente ⅓ de los decesos prematuros (personas menores de 70 años) ocurren debido a ECV. [5]

La mayoría de enfermedades cardiovasculares se pueden prevenir identificando de manera temprana las características conductuales que desencadenan en estas, como, el consumo de tabaco, las dietas poco saludables, la obesidad, la inactividad física y el consumo nocivo de alcohol.

Las personas con presencia de uno o más factores de riesgo que puedan desencadenar en ECV como, hipertensión, diabetes, hiperlipidemia o enfermedades ya establecidas, necesitan de una detección y manejo temprano, en este punto la implementación de una solución de ML que pueda identificar las características que conllevan a un riesgo de ECV y prediga con alta precisión un dictamen sobre estas contribuiría a minimizar los decesos prematuros y no identificados de manera oportuna.

1. PLANTEAMIENTO DE LA SOLUCIÓN

Con el propósito de contribuir a alertar a las personas que puedan estar propensas a sufrir algún tipo de ECV el equipo de trabajo se propuso a desarrollar una solución a través de la cual cualquier persona pudiera obtener una respuesta según sus condiciones patológicas.

A partir de esto se identificaron múltiples soluciones entre las que se recomendó el desarrollo de un modelo de Machine Learning que fuera capaz de aprender las condiciones y características que tienen las personas que están propensas a sufrir una ECV, para lo cual fue necesario desarrollar múltiples pasos que se listan a continuación.

1. Recolección del dataset
2. Definición y entrenamiento del modelo de ML
3. Despliegue del modelo de ML
4. Desarrollo de web para consulta de los usuarios

Teniendo en cuenta estos puntos claves a desarrollar, a continuación se describe de manera detallada cada uno de estos.

**Recolección del dataset**

Para dar una solución a lo planteado anteriormente, se propone inicialmente identificar un dataset existente que tenga el mismo objetivo de clasificación planteado en la problemática, esto para optimizar el tiempo que tomaría obtener una muestra poblacional y generar un conjunto de datos propios. Para esto se utilizó la base de datos de datasets para aplicaciones de inteligencia artificial conocida como Kaggle, y se determinó que se usaría el conjunto de datos “Heart Failure Prediction Dataset”.

Las características de salud que recoge el dataset se muestran a continuación:

* Edad: con el paso de los años aumenta la posibilidad de adquirir distintos factores de riesgo de las ECV, esto especialmente en mayores de 65 años [6],[7].
* Género: datos respaldan que la incidencia de ECV varían dependiendo del género incluso para personas de la misma edad, teniendo una mayor tasa de riesgo los hombres [7],[8].
* Presión arterial: un gran porcentaje de las ECV se atribuyen a una alta presión arterial haciéndola uno de los factores de riesgo más importantes. Es importante tener en cuenta que los valores normales de una persona sana son 120/80 mmHg, mientras que valores superiores implican un riesgo [7],[9].
* Colesterol: niveles elevados de colesterol (hiperlipidemia) están fuertemente relacionados con ECV, por tanto, controlar el colesterol es un método para prevenir enfermedades cardiovasculares. En este parámetro, los niveles normales están en el rango de 125 a 200 mg/dL para personas mayores a 20 años [10],[11].
* Altos niveles de azúcar en la sangre (diabetes): así como la hiperlipidemia, la diabetes también es considerada un factor que aumenta el riesgo de padecer ECV. Para esto se tiene que el rango de una persona sana en ayunas es de 70-110 mg/dL y menor a 140 mg/dL después de comer [12].

**Definición y entrenamiento del modelo de ML**

Posterior a esto, para el entrenamiento del modelo se determina que el tiempo puede reducirse si se lleva a cabo de forma remota y aún más de forma distribuida. Es por esto que se opta por usar recursos de un proveedor de servicios en la nube, en este caso se elige Azure de Microsoft debido a su interfaz es más intuitiva y fácil de usar, sin embargo, también existen otras alternativas que pueden usarse como las proporcionadas por Amazon a través de AWS.

Ahora bien, en Azure se tienen las siguientes herramientas para desarrollar aplicaciones de machine learning:

* Notebooks: se desarrollan los modelos a partir de scripts escritos en algún lenguaje de programación como Python.
* Automated ML: es una herramienta automatizada para optimizar el tiempo y las tareas repetitivas que se suelen realizar cuando se desea implementar un modelo de machine learning. Para esto se definen objetivos y se obtienen modelos que se ajusten a ellos.
* Machine Learning Designer: permite crear modelos de machine learning de forma gráfica a través de bloques predeterminados y configurables que permiten desarrollar la arquitectura deseada.

Para empezar a usar cualquiera de estas herramientas, lo primero que se debe hacer es crear un área de trabajo de machine learning donde se agrupan los recursos que se usan para entrenar y desplegar el modelo. Después, se deben crear los recursos de computación necesarios, una instancia en la que normalmente se desarrolla el modelo, y un cluster de nodos para el entrenamiento distribuido.

**Despliegue del modelo de ML**

Para el despliegue del modelo de ML se hizo uso de la herramienta endpoints que ofrece Azure ML Studio, esta permite desplegar el modelo seleccionado a través de una API, para que pueda ser consumido desde un endpoint con una estructura JSON definida para la realización de las peticiones HTTP y la respuesta al realizar la inferencia en el modelo.

El hacer uso de esta herramienta permitirá que se puedan realizar inferencias del modelo en cualquier otro servicio, permitiendo replicar el trabajo una vez realizado en múltiples soluciones.

**Desarrollo de web para consulta de los usuarios**

Finalmente, con el propósito de permitirle a cualquier tipo de usuario obtener una predicción del modelo de ML entrenado se implementó una aplicación web sencilla en la que se muestran diferentes campos para que el usuario pueda ingresar las características de salud solicitadas y obtenga una respuesta en base a estas.

1. RESULTADOS

En esta sección se muestran los resultados obtenidos de los pasos ejecutados en el planteamiento de la solución, con el propósito de acercar al lector a lo desarrollado por el equipo de trabajo.

**Definición y entrenamiento del modelo de ML**

Para optimizar el tiempo y obtener un modelo que cumpla con el objetivo de la aplicación, se optó por utilizar la herramienta de Automated ML, a partir de la cual se obtuvo múltiples modelos que clasificaron con alta precisión los registros entregados en el dataset, algunos de estos se muestran a continuación.

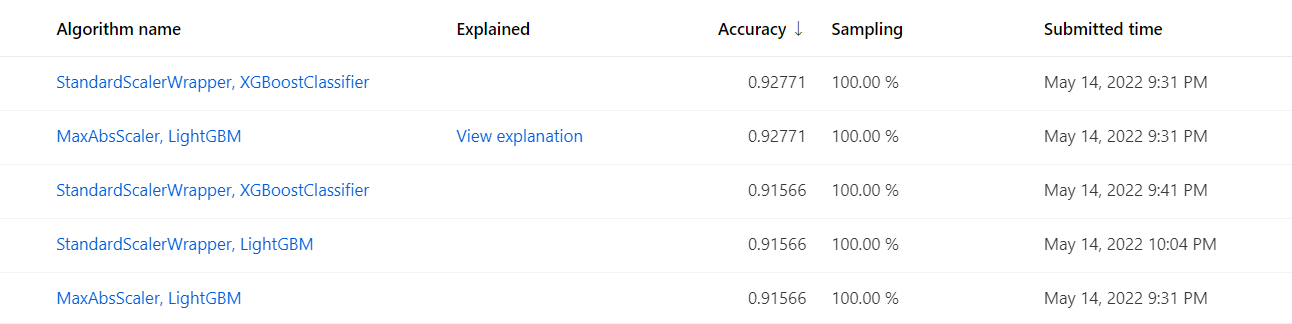


Figura 1. Modelos de ML entrenados en Azure ML Studio.

Al finalizar se determinó que la arquitectura de modelo que mejor funciona es StandardScalerWrapper, XGBoostClassifier con una precisión en la predicción del 92,771%para los datos de validación asignados para el modelo.

**Despliegue del modelo de ML**

Posterior a la selección del modelo de ML que mejores resultados demostró para la clasificación de fallos cardiacos haciendo empleo del dataset seleccionado, se procedió a desplegar este modelo en la web para que fuera accesible y consumible desde otros servicios.

Al hacer uso de la herramienta endpoints de Azure se obtuvo la siguiente estructura JSON para la realización de las peticiones POST al modelo de ML.

*{*

*"Inputs": {*

*"data": [*

*{*

*"Age": 0,*

*"Sex": "valor",*

*"ChestPainType": "valor",*

*"RestingBP": 0,*

*"Cholesterol": 0,*

*"FastingBS": 0,*

*"RestingECG": "valor",*

*"MaxHR": 0,*

*"ExerciseAngina": "valor",*

*"Oldpeak": 0,*

*"ST\_Slope": "valor"*

*}*

*]*

*},*

*"GlobalParameters": {*

*"method": "predict"*

*}*

*}*

De la misma manera se obtuvo la siguiente estructura JSON para las respuestas entregadas por el modelo posterior a la petición POST.

*{*

*"Results": [*

*0*

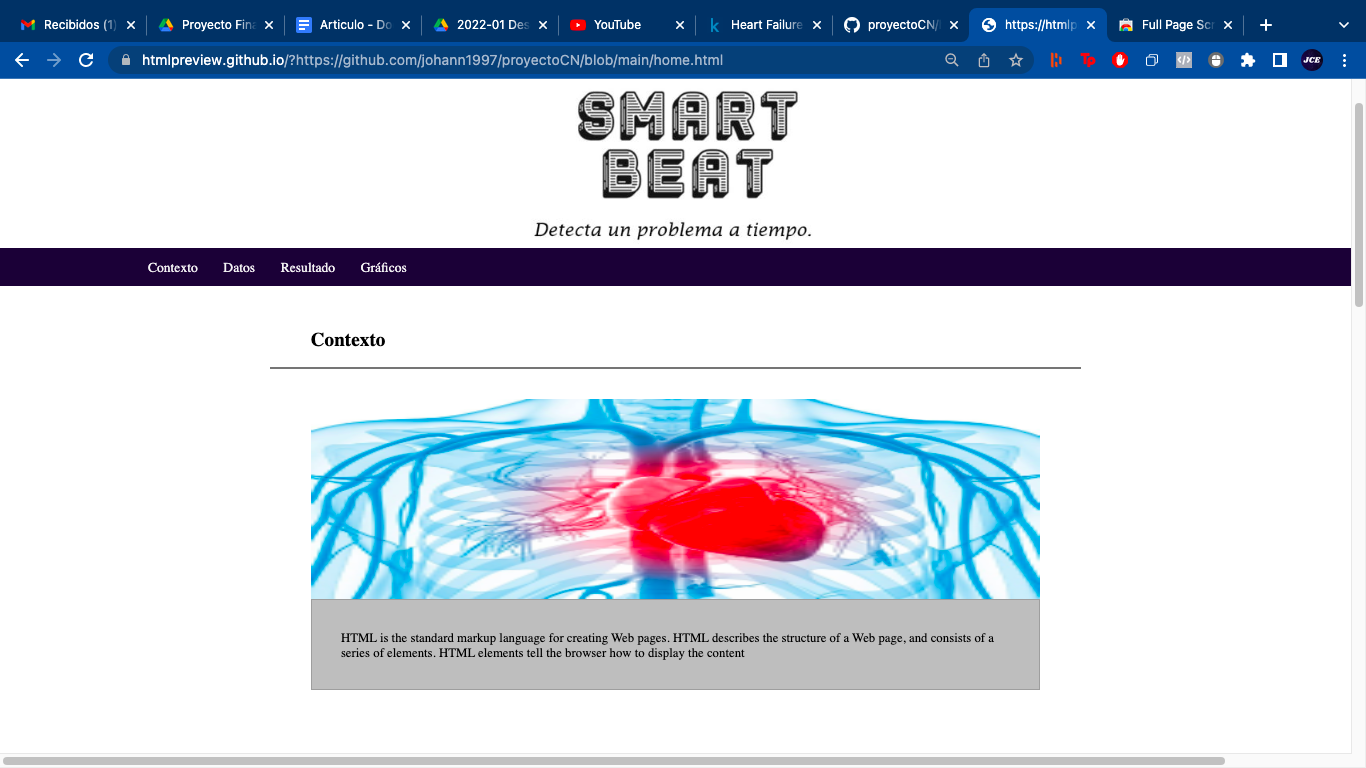
*]*

*}*

Ambas estructuras fueron vitales para el correcto desarrollo de la web con la que el usuario podrá hacer uso del modelo de ML.

**Desarrollo de web para consulta de los usuarios**

Finalmente se desarrolló la web en la que se permitiría a los usuarios hacer uso del modelo, esta es una web sencilla que ofrece un contexto general de la problemática abordada, muestra las entradas para los datos de los usuarios y un campo de salida para la respuesta obtenida del modelo, a continuación se muestra la web desarrollada.



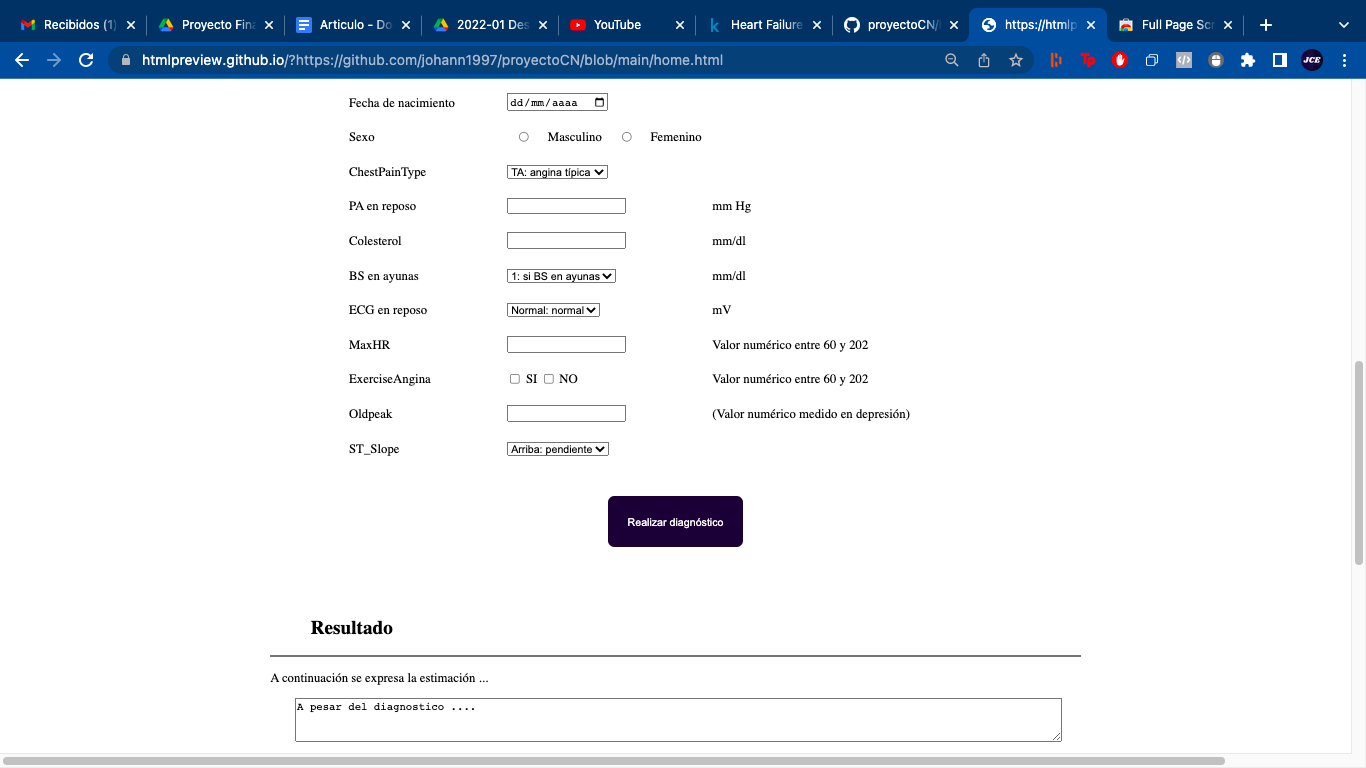


Figura 2. Web desarrollada para la consulta de los usuarios.

1. CONCLUSIONES

Cuando se realizan aplicaciones de machine learning es muy importante tener diferentes modelos con el objetivo de evaluar cuál se adapta mejor a los requerimientos. Para esto es clave hacer uso de una herramienta como Automated ML con la que se puede explorar la respuesta de una gran cantidad de modelos de forma automatizada, lo que ahorra tiempo al desarrollador y le permite enfocarse en los casos de uso de la aplicación resultando en un mejor desempeño de esta.

Los recursos de desarrollo para aplicaciones de machine learning ofrecidos por azure son una gran oportunidad de trabajo, debido a que no se limitan a un nivel de conocimiento avanzado. Por lo contrario, sus distintas herramientas permiten que diferentes personas puedan implementar la aplicación que deseen. Aunque es evidente que a un mayor nivel de experiencia se tiene un mayor control sobre el desarrollo, lo cual es similar a los modelos de servicio de computación en la nube.

Tecnologías como el Machine Learning ofrecen la posibilidad de desarrollar aplicaciones que contribuyan a solucionar problemas muy complejos en los que el procesamiento y análisis de datos sea un punto clave, para esto es importante centrarse en identificar y/o recolectar conjuntos de datos cualificados, que permitan obtener soluciones efectivas para un grupo de usuarios específicos.

1. REFERENCIAS

[1] R. Fahim El-Gazzar, “IFIP AICT 429 - A Literature Review on Cloud Computing Adoption Issues in Enterprises,” pp. 214–242, 2014.

[2] K. Y. Ngiam and I. W. Khor, “Big data and machine learning algorithms for health-care delivery,” Lancet Oncol., vol. 20, no. 5, pp. e262–e273, May 2019, doi: 10.1016/S1470-2045(19)30149-4.

[3] G. Edwards, “Machine Learning | An Introduction.” https://towardsdatascience.com/machine-learning-an-introduction-23b84d51e6d0 (accessed May 17, 2022).

[4] Diapositivas del curso

[5]“Cardiovascular diseases (CVDs).” https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (accessed May 14, 2022).

[6] “Factores de riesgo cardiovascular | Texas Heart Institute.” https://www.texasheart.org/heart-health/heart-information-center/topics/factores-de-riesgo-cardiovascular/ (accessed May 14, 2022).

[7] R. Dhingra and R. S. Vasan, “Age as a Cardiovascular Risk Factor,” The Medical Clinics of North America, vol. 96, no. 1, p. 87, Jan. 2012, doi: 10.1016/J.MCNA.2011.11.003.

[8] I. Wakabayashi, “Gender differences in cardiovascular risk factors in patients with coronary artery disease and those with type 2 diabetes,” Journal of Thoracic Disease, vol. 9, no. 5, p. E503, May 2017, doi: 10.21037/JTD.2017.04.30.

[9] C. Y. Wu, H. Y. Hu, Y. J. Chou, N. Huang, Y. C. Chou, and C. P. Li, “High Blood Pressure and All-Cause and Cardiovascular Disease Mortalities in Community-Dwelling Older Adults,” Medicine, vol. 94, no. 47, p. e2160, Nov. 2015, doi: 10.1097/MD.0000000000002160.

[10] “Niveles de colesterol: Lo que usted debe saber: MedlinePlus en español.” https://medlineplus.gov/spanish/cholesterollevelswhatyouneedtoknow.html (accessed May 14, 2022).

[11] R. H. Nelson, “Hyperlipidemia as a Risk Factor for Cardiovascular Disease,” Prim Care, vol. 40, no. 1, p. 195, Mar. 2013, doi: 10.1016/J.POP.2012.11.003.

[12] C. Park et al., “Fasting glucose level and the risk of incident atherosclerotic cardiovascular diseases,” Diabetes Care, vol. 36, no. 7, pp. 1988–1993, 2013, doi: 10.2337/DC12-1577/-/DC1.