



Universidade Federal de Pernambuco



EDITAL FACEPE Nº 18/2024
APQ – Universal (Auxílio a Projetos de Pesquisa)

FAIXA A

Identificação:

Título da proposta	Análise geométrica e topologia espectral de marcadores fisiológicos para indicadores pré-clínicos de doenças cardiovasculares e síndrome metabólica em jovens e adultos
Área	Ciências da Saúde
Tema	Saúde Humana
Coordenadora Geral	Carol Virgínia Góis Leandro - Professora Titular do Núcleo de Nutrição do Centro Acadêmico de Vitória de Santo Antão – UFPE. Bolsista de Produtividade PQ 1D do CNPq. Email: carol.leandro@ufpe.br
Equipe	Cleber Zanchettin . Professor Associado do Centro de Informática da UFPE (Cin-UFPE). Bolsista de Produtividade Desen. Tec. e Extensão Inovadora do CNPq - Nível 2. Email: cz@cin.ufpe.br
	Adriano Augusto de Moraes Sarmento – Professor Associado do Centro de Informática da UFPE (Cin – UFPE). Email: aams@cin.ufpe.br
	Wylla Tatiana Ferreira e Silva – Professora Associada do Núcleo de Nutrição do CAV-UFPE. Email: wylla.silva@ufpe.br
Pesquisadores/as convidados/as	Sônia Maria Oliveira Cavalcanti Marinho – Professora Associada do Núcleo de Nutrição do CAV-UFPE. Email: sonia.marinho@ufpe.br
	Luciano Pirola - CarMeN (Cardiology, Metabolism and Nutrition) Laboratory, INSERM U1060, South Lyon Medical Faculty, Lyon-1 University, Oullins, France. Luciano.pirola@univ-lyon1.fr
	Jaime Cesar dos Santos Filho - Professor Adjunto do Departamento de Matemática da Unidade Acadêmica do Cabo de Santo Agostinho – UFRPE. Email: jaime.cesar@ufrpe.br
	André Luiz de Góes Pacheco – Pós-doutorado pelo Programa de Pós-graduação em Biologia Aplicada à Saúde da UFPE. Email: algp@cin.ufpe.br
	Isabele Goes Nobre – Pós-doutorado pelo Programa de Pós-graduação em Nutrição (PPGNutri – UFPE). Email: isabele.gnobre@ufpe.br
	Gabriela Carvalho Jurema Santos – Doutora pelo Programa de Pós-graduação em Nutrição (PPGNutri – UFPE). Email: gabriela.cjsantos@ufpe.br
Doutoranda	Élida Karine de Lira Ferreira – Doutoranda pelo Programa de Pós-graduação em Biologia Aplicada à Saúde da UFPE. Email: eklf@cin.ufpe.br
Mestrandos/as	Rayssa Franciely Temudo dos Santos – Mestranda pelo Programa de Pós-graduação em Nutrição, Atividade Física e Plasticidade Fenotípica do CAV-UFPE (PPGNAFPF-UFPE). Email: rfts@cin.ufpe.br
	Denise Mirelli Leão Costa – Mestranda pelo Programa de Pós-graduação em Nutrição, Atividade Física e Plasticidade Fenotípica do CAV-UFPE (PPGNAFPF-UFPE). Email: denise.leao@ufpe.br

	Willian Motta Bocanegra – Mestrando pelo Programa de Pós-graduação em Ciências de Materiais pela UFPE. Email: willian.bocanegra@ufpe.br
	Patrícia Miller Simões – Mestranda pelo Programa de Pós-graduação em Nutrição, Atividade Física e Plasticidade Fenotípica do CAV-UPFE (PPGNAFPF-UFPE). Email: patricia.millersimoes@ufpe.br
	Fabiane de Oliveira Freitas – Mestranda pelo Programa de Pós-graduação em Nutrição, Atividade Física e Plasticidade Fenotípica do CAV-UPFE (PPGNAFPF-UFPE). Email: fabiane.freitas@ufpe.br
Instituição da Coordenadora Geral	Centro Acadêmico de Vitória – Universidade Federal de Pernambuco
Resumo	O presente projeto tem como objetivo desenvolver análise geométrica e topologia espectral de marcadores fisiológicos, bioquímicos e comportamentais para indicadores pré-clínicos de doenças cardiovasculares e síndrome metabólica em jovens e adultos residentes na Zona da Mata do estado de Pernambuco. Serão avaliados 300 indivíduos de ambos os gêneros (idade entre 25 e 55 anos) residentes da Zona da Mata Sul de Pernambuco e que estão sendo acompanhados no Hospital das Clínicas/EBSERH da UFPE. Serão realizados dois estudos: No estudo 1, serão avaliados 100 pacientes do setor de Cardiologia e Endocrinologia do HC UFPE/EBSERH divididos em 2 grupos: pré-Diabetes Mellitus tipo 2 (preDMT2, n = 50) e DMT2 (n = 50). O grupo normoglicêmico/controle (n=50) será constituído de indivíduos não-pacientes do HC-UFPE/EBSERH. No estudo 2, serão avaliados marcadores de pré-hipertensão em 100 pacientes do setor de Cardiologia do HC UFPE/EBSERH. O grupo normotenso/controle (n=50) será constituído de indivíduos não-pacientes do HC-UFPE/EBSERH. Serão avaliados: antropometria e composição corporal (cluster morfológico), nível de atividade física habitual e consumo alimentar (cluster comportamental), nível salivar de cortisol e glicemia de jejum (cluster bioquímico) e variabilidade cardíaca, frequência cardíaca, consumo máximo de oxigênio e pressão arterial (cluster fisiológico) Pretende-se desenvolver três tipos de redes complexas multidimensionais para cada indivíduo, e para os grupos, permitindo uma visão espectral individual e global. Espera-se identificar as diferenças nos espectros entre esses conjuntos para compreender como a geometria e topologia entre os grupos mudam, permitindo caracterizar um indivíduo unicamente a partir do seu espectro e saber o quão distante ele está dos grupos com comorbidades. Com estes resultados, iremos construir objetos multidimensionais formados por triplas de variáveis fortemente correlacionadas em ambas as redes complexas desenvolvidas, que chamaremos de "tripleτες", que podem ser usados dinamicamente para saber se nos aproximamos ou distanciamos dos potenciais indicadores de doenças cardiovasculares e metabólicas. Os dados reais irão subsidiar a criação de conjuntos de dados sintéticos mais amplos e diversificados para treinamento e validação dos modelos de Aprendizado de Máquina (<i>machine learning</i>) e Aprendizado Profundo (<i>Deep Learning</i>). Os resultados gerados nesse estudo, comporão um banco de dados anonimizado com o objetivo de serem utilizados para o desenvolvimento e melhoramento de algoritmos, pesquisa, formação de recursos humanos e publicações científicas.
Palavras-chave:	Hipertensão Arterial; Diabetes Mellitus Tipo 2; Geometria; <i>Machine Learning</i> ; <i>Deep Learning</i> ; Inteligência Artificial; Consumo Alimentar

2 DETALHAMENTO

2.1 Introdução

De acordo com a Organização Mundial da Saúde (OMS), as doenças cardiovasculares (DCV) são a principal causa de morte no mundo, sendo responsáveis por quase 18 milhões de mortes anualmente. Isso representa aproximadamente 31% de todas as mortes no mundo (<https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death> acessado em 21 de junho de 2024). No Brasil, a prevalência estimada de diabetes tipo 2 é de 9,2%, enquanto a prevalência de hipertensão na população acima de 18 anos é de 23,9% e de obesidade alcançou 30,2% (MAYORAL; ANDRADE; MAYORAL; HUERTA *et al.*, 2020; MUZY; CAMPOS; EMMERICK; SILVA *et al.*, 2021). De acordo com a Pesquisa Nacional de Saúde (PNS, 2019), a prevalência de hipertensão arterial na população residente no Nordeste do Brasil é de 22,5%, especificamente em Pernambuco, a prevalência atinge 24% (FEITOSA; BARROSO; MION JUNIOR; NOBRE *et al.*, 2024; JULIAO; SOUZA; GUIMARAES, 2021).

A diabetes mellitus tipo 2 (T2DM) é uma doença endócrina de maior prevalência em todo o mundo, caracterizada por dislipidemia e disglucemia. Segundo a **Federação Internacional de Diabetes (IDF)**, 537 milhões de adultos foram diagnosticados com diabetes tipo 2, e três em cada quatro adultos com diabetes vivem em países de baixa e média-renda (<https://idf.org/about-diabetes/what-is-diabetes/>, acessado em 21 de junho de 2024). Ainda segundo a IDF, estima-se que 700,2 milhões de adultos com idades entre 20 e 79 anos sofrerão de DM2 até 2045. A pré-diabetes mellitus (préDM) é um estado de disglucemia que precede o início e o desenvolvimento do DM2 (MONIKA; ARORA; KUMAR; POPLI, 2024). Atualmente, a prevalência de PreDM em adultos é relatado como sendo de 38% nos EUA e 25,7% no Brasil (FEITOSA; BARROSO; MION JUNIOR; NOBRE *et al.*, 2024). A T2DM e a PreDM se tornaram um dos principais desafios da saúde pública no mundo. A pré-diabetes aumenta o risco de diabetes tipo 2, síndrome metabólica, doença renal crônica e doença cardiovascular. Da mesma forma, a hipertensão arterial é um grave problema para a saúde global e uma das principais causas de morte prematura em todo o mundo, afeta aproximadamente mil milhões de pessoas em todo o mundo (MONIKA; ARORA; KUMAR; POPLI, 2024).

A pré-hipertensão, um estágio intermediário entre a pressão arterial normal e a alta, afeta 25 a 50% dos adultos em todo o mundo (KANAZAWA; SUGIMOTO, 2018). Reconhecida como precursora da hipertensão, a pré-hipertensão pode converter-se em hipertensão a uma taxa de 30% num período de quatro anos e contribuir para a progressão da doença cardiovascular (SRIVASTAVA; MIRZA; VAQAR; SHARAN, 2024). Fatores de risco cardiovascular, incluindo genética, tabagismo, obesidade e inatividade física estão ligados à pré-hipertensão (ENDRIAS; TESFAYE MAMITO; GETA HARDIDO; ATARO, 2024). A detecção precoce da pré-hipertensão pode prevenir a hipertensão e o risco de doenças cardiovasculares (SRIVASTAVA; MIRZA; VAQAR; SHARAN, 2024). As intervenções não farmacológicas para o manejo da pré-hipertensão incluem dieta, exercícios, cessação do tabagismo, redução da ingestão de sal e álcool, controle do peso corporal e redução do estresse (CHARCHAR; PRESTES; MILLS; CHING *et al.*, 2024;

FEITOSA; BARROSO; MION JUNIOR; NOBRE *et al.*, 2024; SRIVASTAVA; MIRZA; VAQAR; SHARAN, 2024). Estudos clínicos demonstram que a identificação precoce de marcadores bioquímicos, fisiológicos, morfológico (antropometria e composição corporal) podem impedir o indivíduo de progredir com a instalação da DCV propriamente dita (NAKANGA; CRAMPIN; MKANDAWIRE; BANDA *et al.*, 2023; OBRYCKI; SARNECKI; PAC; DEREZINSKI *et al.*, 2023).

Em um estudo realizado com 481 indivíduos adultos chineses divididos em 3 grupos (normoglicêmicos, preDM e DMT2), mostrou que biomarcadores lipídicos foram identificados em indivíduos preDM e que podem ser utilizados no diagnóstico clínico quando controlados por sexo, idade e índice de massa corporal (IMC) (XUAN; HU; ZHANG; WANG *et al.*, 2022). Em um estudo realizado no Malawi, foram avaliados 389 indivíduos (idade 48 ± 15 anos) preDM que foram acompanhados durante 4 anos e 26% progrediram para DMT2 (NAKANGA; CRAMPIN; MKANDAWIRE; BANDA *et al.*, 2023). Os preditores de progressão foram: idade (odds ratio [OR] 1,48, $P = 0,046$), IMC (OR 1,98, $P = 0,001$), circunferência da cintura (OR 2,50, $P < 0,001$), relação cintura-quadril (OR 1,40, $P = 0,03$), pressão arterial sistólica (OR 1,56, $P = 0,01$), glicemia de jejum (OR 1,53, $P = 0,01$), colesterol (OR 1,44, $P = 0,05$) e frações de colesterol de baixa densidade (OR 1,80, $P = 0,002$) (NAKANGA; CRAMPIN; MKANDAWIRE; BANDA *et al.*, 2023). Um modelo simples combinando glicemia de jejum e circunferência da cintura foi preditivo de progressão para diabetes (área ROC sob a curva = 0,79).

Um estudo recente identificou seis biomarcadores metabólicos em indivíduos pré-hipertensos ($n = 98$, pressão arterial 120–139 mmHg), hipertensos ($n=72$, pressão arterial ≥ 140 mmHg) e em indivíduos saudáveis ($n = 72$, pressão arterial < 120 mmHg) que podem distinguir a pré-hipertensão da fisiopatologia da hipertensão (AL ASHMAR; ANWARDEEN; ANLAR; PEDERSEN *et al.*, 2024). Houve correlação do aparecimento destes metabólitos com o IMC e com o perfil lipídico dos indivíduos (AL ASHMAR; ANWARDEEN; ANLAR; PEDERSEN *et al.*, 2024).

O desenvolvimento de ferramentas que possam capturar sinais biológicos e correlacionar com biomarcadores bioquímicos e fisiológicos para acompanhar a progressão de um estado pré-diabético ou de pré-hipertensão é necessário. Há uma forte relação entre indicadores moleculares e celulares do aparecimento precoce de DCV e variações morfológicas (massa gorda, massa magra, circunferência da cintura, relação cintura/quadril, percentual de gordura corporal, índice de massa corporal), comportamental (consumo alimentar e atividade física regular), fisiológicas (temperatura corporal, variabilidade cardíaca, pressão arterial sistólica e diastólica, consumo máximo de oxigênio e saturação de oxigênio) e bioquímica (níveis salivares de cortisol, resistência à insulina e glicemia de jejum) dos indivíduos.

Para o desenvolvimento eficaz dessas ferramentas, é crucial ter acesso a uma quantidade substancial de dados. Coletar dados diretamente de um grupo de indivíduos é um processo dispendioso e repleto de desafios, como questões de privacidade, direitos de uso e disponibilidade dos voluntários para participar da coleta (SINGH; KUMAR; KHANNA; LAIRD *et al.*, 2024). Considerando que a criação de uma ferramenta baseada em Inteligência Artificial para diagnosticar doenças cardiovasculares requer uma quantidade significativa de amostras (considerando diversas situações, casos e protocolos), o uso de técnicas de aumento de dados, através da geração de dados sintéticos, surge como uma alternativa viável

e menos onerosa, capaz de mitigar tais dificuldades e permitir a criação de modelos robustos, confiáveis e interpretáveis (SINGH; KUMAR; KHANNA; LAIRD *et al.*, 2024).

Estudos têm demonstrado que o aprendizado profundo (*deep learning*) é eficaz na resolução de diversos problemas complexos, inclusive na área da saúde, na qual técnicas como Redes Neurais Convolucionais (CNNs), Redes *Long Short-term Memory* (LSTM), entre outras, têm se destacado (RAO; LU; WU; ZHAO *et al.*, 2024; TROTTET; ALLAM; HORVATH; FINCKH *et al.*, 2024). Neste contexto, uma aplicação relevante dessas redes é a geração de dados sintéticos para sensores utilizados em soluções de saúde. Recentemente, foram propostos métodos com desempenho superior ao estado-da-arte na geração de dados sintéticos para detecção de diabetes, empregando *autoencoders* compostos por redes CNNs (PARANJPE; WANG; ANANDAKRISHNAN; HAYDAK *et al.*, 2023; VYAS; DESHPANDE; SEN; KIM *et al.*, 2023), bem como para o aumento de dados de acelerômetros para o monitoramento da doença de Parkinson (MARTINEZ-MURCIA; ARCO; JIMENEZ-MESA; SEGOVIA *et al.*, 2024).

A utilização de dados sintéticos desempenha um papel fundamental nesse processo, permitindo a criação de conjuntos de dados mais amplos e diversificados para treinamento e validação dos modelos de Aprendizado de Máquina (*machine learning*) e *deep learning* (NANDAGOPAL; SEERANGAN; GOVINDARAJU; ABI *et al.*, 2024). Ao gerar dados sintéticos para aumentar os dados coletados, se pode enriquecer o conjunto de dados disponíveis, tornando-o mais representativo em relação à diversidade da população. Este modelo tem sido utilizado em estudos prévios com doenças cardiovasculares, neurológicas e inflamatórias (MARTINEZ-MURCIA; ARCO; JIMENEZ-MESA; SEGOVIA *et al.*, 2024; NANDAGOPAL; SEERANGAN; GOVINDARAJU; ABI *et al.*, 2024; RAO; LU; WU; ZHAO *et al.*, 2024; SINGH; KUMAR; KHANNA; LAIRD *et al.*, 2024; TROTTET; ALLAM; HORVATH; FINCKH *et al.*, 2024).

Objetivos:

Geral: Desenvolver clusters de indicadores morfológicos, comportamentais, bioquímicos e fisiológicos para o diagnóstico pré-clínico de hipertensão arterial e diabetes mellitus não-insulino dependente (Figura 1). Os sinais biológicos serão capturados por dispositivo e analisados por geometria e topologia espectral.

Específicos:

- Desenvolver três tipos de redes complexas multidimensionais para cada indivíduo em nossa coleta, além de redes complexas para os grupos, permitindo uma visão espectral individual e global.
- Identificar as diferenças nos espectros entre esses conjuntos para compreender como a geometria e topologia entre os grupos mudam, permitindo caracterizar um indivíduo unicamente a partir do seu espectro e saber o quão distante ele está dos grupos com comorbidades.
- Construir objetos multidimensionais formados por triplas de variáveis fortemente correlacionadas em ambas as redes complexas desenvolvidas, que chamaremos de "tripletes", que podem ser usados dinamicamente para saber se nos aproximamos ou distanciamos dos potenciais indicadores das doenças cardiovasculares (DCV).

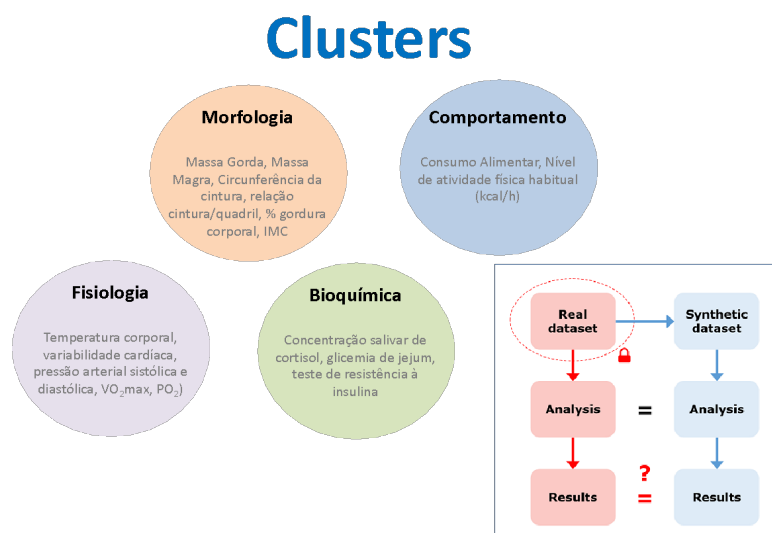


Figura 1. Diferentes clusters que serão utilizados como potenciais biomarcadores do aparecimento precoce de hipertensão arterial e diabetes mellitus tipo 2.

Hipótese

A captura de sinais biológicos (temperatura corporal basal e após o esforço, variabilidade cardíaca basal e após esforço, concentração salivar de cortisol, composição corporal e antropometria, consumo alimentar, nível de atividade física habitual, pressão arterial sistólica e diastólica e saturação de O_2 basal e no esforço) são marcadores dinâmicos de quadros pré-clínicos de doenças cardiovasculares (obesidade, dislipidemia, diabetes tipo 2 e hipertensão arterial).

2.2 Produção Científica e Tecnológica

Lista dos Lista dos cinco produtos mais importantes de autoria da coordenadora nos últimos 5 anos

1. Jéssica de Oliveira Campos, Tafnes Laís Pereira Santos de Almeida Oliveira, Oriane Vitalis, Jéssica Gonzaga Pereira, Isabella da Costa Ribeiro Nogueira, Gabriela Carvalho Jurema Santos, Karim Chikh, **Carol Gois Leandro**, João Henrique da Costa-Silva, Luciano Pirola. Association between Childhood Overweight and Altered Concentrations of Circulating Amino Acids. *Nutrients*. 2024 Jun 12;16(12):1843. doi: 10.3390/nu16121843.
2. Eulálio Malinga, **Carol Góis Leandro**, Fernanda Trigueiro de Almeida Araujo, Rafael Dos Santos Henrique, Mario Eugénio Tchamo, Wylla Tatiana Ferreira e Silva. Birth weight and nutritional status in school-age children from Boane city, Mozambique. *Am J Hum Biol*. 2024 Mar 19:e24072. doi: 10.1002/ajhb.24072. Online ahead of print.
3. **Carol Góis Leandro**, Conrado Rodrigues, Pedro Carelli, Charles Morphy D Santos. Rebuilding an agenda for Brazilian science and technology. *Lancet*. 2023 Feb 25;401(10377):642-643. doi: 10.1016/S0140-6736(23)00098-3.

4. Isabella da Costa Ribeiro, Tafnes Laís Pereira Santos de Almeida Oliveira, Gabriela Carvalho Jurema Santos, Isabele Góes Nobre, Ravi Marinho Dos Santos, Camilla Peixoto Santos Rodrigues, Maria Vitória Dos Santos Costa, Maria Carla Melo Damasceno, Wylla Tatiana Ferreira e Silva, **Carol Góis Leandro**. Daily consumption of ultra-processed foods and cardiometabolic risk factors in children aged 7 to 10 years in Northeast Brazil. *Nutr Health*. 2023 Sep;29(3):557-565. doi: 10.1177/02601060221084816. Epub 2022 Mar 7.
5. Isabele Góes Nobre, Gabriela Carvalho Jurema Santos, Tafnes Laís Pereira Santos de Almeida Oliveira, Isabella da Costa Ribeiro, Ravi Marinho Dos Santos, Camilla Peixoto Santos Rodrigues, Marcos André Moura-Dos-Santos, Julie-Anne Nazare, Luciano Pirola, **Carol Góis Leandro**. Food consumption habits, gestational age and birth weight are predictive for children with excess weight: An analysis based on artificial neural network. *Nutr Health*. 2022 Sep 16:2601060221124040. doi: 10.1177/02601060221124040. Online ahead of print.

2.3. Métodos

Aspectos éticos

O presente projeto foi submetido ao Comitê de Ética em Pesquisa (CAAE: 81126324.8.0000.5208) para estudos com humanos da Universidade Federal de Pernambuco para análise. Uma vez que serão utilizados dados de pacientes com diagnóstico pré-clínico e clínico de diabetes mellitus tipo II e hipertensão arterial, o projeto também foi submetido e já está aprovado pelo Comitê de Ética em Pesquisa (CEP) do Hospital das Clínicas UFPE/EBSERH.

Desenho do estudo

Uma amostra de 300 indivíduos (com idade entre os 25 e 55 anos) de ambos os gêneros será submetida a avaliação morfológica, comportamental, bioquímica e fisiológica. Os dados reais serão amplificados em base sintética.

O **estudo 1** irá avaliar marcadores de pré-diabetes de pacientes do setor de Cardiologia e Endocrinologia do Hospital das Clínicas da Universidade Federal de Pernambuco/Empresa Brasileira de Serviços Hospitalares do Ministério da Educação (UFPE/EBSERH do MEC). Uma subamostra de 150 indivíduos será dividida em 3 grupos: normoglicêmicos ($n = 50$), preDM ($n = 50$) e DMT2 ($n = 50$). Os critérios diagnósticos serão aplicados de acordo com os critérios diagnósticos da Organização Mundial da Saúde (OMS 1999). PreDM é dividido em três subcategorias diferentes: deficiente tolerância à glicose (IGT) com $7,8 \text{ mmol/L} \leq 2\text{h-PG} < 11,1$, glicemia de jejum (IFG) com $6,1 \text{ mmol/L} \leq \text{FPG} < 7,0$ e IGT&IFG com $7,8 \text{ mmol/L} \leq 2\text{h-PG} < 11,1$ e $6,1 \text{ mmol/L} \leq \text{FPG} < 7,1$ com base em um teste oral de tolerância à glicose (OGTT).

O **estudo 2** irá avaliar marcadores de pré-hipertensão em 150 pacientes do setor de Cardiologia do HC UFPE/EBSERH. Os grupos de estudo serão compostos por participantes hipertensos (pressão arterial $\geq 140 \text{ mmHg}$, $n = 50$), pré-hipertensos (pressão arterial 120–139 mmHg, $n = 50$) e um grupo controle de participantes normotensos (pressão arterial $< 120 \text{ mmHg}$, $n = 50$).

População, cálculo amostral e critérios de inclusão e exclusão

No total, serão avaliados 300 indivíduos (com idade entre os 25 e 55 anos) de ambos os gêneros, cadastrados no HC UFPE/EBSERH. A amostragem será por conveniência. Para cada participante, serão

realizadas medidas antropométricas como altura, peso, circunferência da cintura e quadril, composição corporal e estado nutricional. Também será realizada avaliação eletrocardiográfica e pressão arterial. Os dados de parâmetros bioquímicos incluindo hemograma completo, perfil lipídico, HBA1C, testes de função hepática e proteína C reativa serão obtidos através do prontuário dos pacientes. Os dados reais serão amplificados em base sintética.

Poderão participar do presente estudo indivíduos adultos ≥ 25 anos de idade, de ambos os gêneros com diagnóstico pré-clínico e clínico de doença cardiovascular (diabetes tipo 2 e hipertensão), podendo apresentar comorbidades como dislipidemia ou sobrepeso e obesidade. Para compor o grupo controle, poderão participar indivíduos adultos ≥ 25 anos de idade, de ambos os gêneros sem diagnóstico de doença cardiovascular. Indivíduos que estejam em internamento ou na fase aguda da doença, ou apresentarem algum déficit motor ou que se recusem a assinar o termo de consentimento livre e esclarecido não poderão participar do estudo.

Visitas/ consultas

A coleta de dados será realizada no setor de cardiologia do HC-UFPE/EBSERH nos turnos da manhã e da tarde a partir do mês de Outubro/2024. Os pacientes serão avaliados em balança de bioimpedância e com um eletrocardiograma. Os dados referentes às avaliações bioquímicas serão obtidos no prontuário do próprio paciente ao longo do período de acompanhamento e de evolução da doença. Após a avaliação fisiológica e de composição corporal, os pacientes receberão um questionário sobre os hábitos alimentares e o nível de atividade física diário.

Variáveis

Antropometria e Composição corporal:

Para cada participante, medidas antropométricas como altura, peso, Índice de Massa Corporal (IMC), razão cintura e quadril, total de água corporal, conteúdo de proteína e minerais, massa livre de gordura, massa muscular, percentual de gordura corporal, massa gorda, percentual de gordura segmentar (braços, tronco e pernas) e taxa metabólica basal serão obtidas usando uma balança de BioImpedância InBody (Ottoboni, Rio de Janeiro, Brasil).

Questionário de Frequência Alimentar (QFA)

Os dados de consumo alimentar serão coletados através de questionário de frequência alimentar e recordatório 24h. Após os resultados do QFA, aplicaremos o questionário de consumo alimentar aos pacientes. Link para acesso ao Questionário de Frequência Alimentar de Adultos: https://gac-usp.com.br/wp-content/uploads/2020/01/QFA_Adulto.pdf

Questionário de Atividade Física (IPAQ)

O Questionário Internacional de Atividade Física - IPAQ estima o dispêndio energético semanal de atividades físicas (AF). Todos os indivíduos serão orientados a preencher o questionário durante o período de avaliação. Link para acesso ao Questionário Internacional de Atividade Física para Adultos: https://www.uel.br/grupo-pesquisa/gepafe/acesso-restrito/Question%E1rios/Ipaq_versao_curta_questionario.pdf

Registro de eletrocardiograma, análise da variabilidade da frequência cardíaca e atividade elétrica cardíaca

A medida eletrocardiográfica será registrada pelo equipamento Equivital Wearable ECG Starter Pack ADInstruments® PowerLab® (modelo 26T-LTS) e os registros realizados com a configuração de 5 eletrodos utilizando o software de aquisição de dados LabChart (ADInstruments®, Bella Vista, NSW, Austrália). Para essas análises, o ECG será configurado da seguinte forma: 80 milissegundos para determinar a largura típica do QRS, 300 milissegundos para determinar a distância entre as ondas R, 240 milissegundos para os intervalos PR e TR máximos.

Etapas de validação das variáveis antropométricas, comportamentais, fisiológicas e bioquímicas:

No total, 10% da amostra será submetida a um teste de validação do instrumento de captura de sinais. Os dados antropométricos coletados através de bioimpedância serão validados pelo DEXA (densitometria por emissão de raios-X de dupla energia). Os dados fisiológicos: glicemia, níveis de oxigênio no sangue (SpO2), pressão arterial, ECG, índice diferencial de VFC, frequência cardíaca, recuperação da frequência cardíaca, variabilidade da frequência cardíaca (VFC), serão validados em eletrocardiograma. Os dados de atividade física habitual e gasto energético serão validados por Ergoespirometria em laboratório. Os dados de consumo alimentar serão validados pela reaplicação do questionário.

Análise dos dados

Utilizando os dados de cada cluster, transformaremos cada indivíduo em um *dataframe*. Esse processo nos possibilita extrair *insights* globais sobre os aspectos geométricos e topológicos envolvidos. Com base nas matrizes, desenvolveremos grafos, hipergrafos e complexos simpliciais tanto para indivíduos quanto para grupos, permitindo a análise de centralidade em grafos e hipergrafos, além de características relevantes à Análise Topológica de Dados (TDA), como entropias e filtrações. Embora os hipergrafos e os complexos simpliciais sejam estruturas matemáticas distintas, eles compartilham certas similaridades, especialmente no que diz respeito à relação entre k -hiperarestas em hipergrafos e $(k-1)$ -simplexos em complexos simpliciais. Isso possibilita a comparação de objetos de dimensão k nessas estruturas. Especificamente para $k = 3$, analisamos tripletes $[var1, var2, var3]$ em ambas as estruturas, identificando aqueles com maior relevância em termos de centralidade nas redes de hipergrafos e de robustez nas filtrações dos complexos simpliciais. A figura 2 apresenta um exemplo das centralidades multidimensionais dos hipergrafos (eixo x) e alguns tripletes (eixo y).

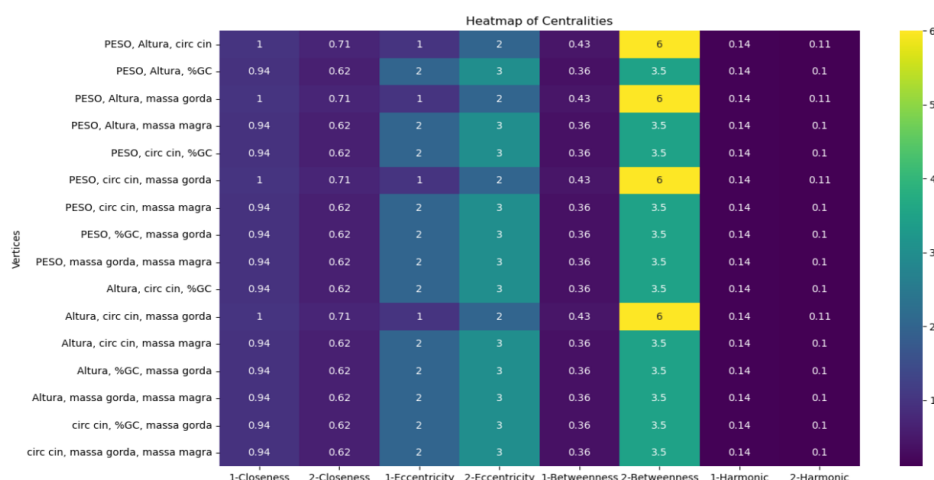


Figura 2. Exemplo das centralidades multidimensionais dos hipergrafos (eixo x) e tripletes (eixo y). Elaboração própria utilizando bases de dados de estudos prévios do grupo de pesquisa.

Treinamento de dados

Iremos utilizar o protocolo MIXED (treinar com dados sintéticos e reais misturados e testar com dados reais) avalia a capacidade dos dados sintéticos de ampliar o conjunto de dados original, visando aumentar a eficiência de um classificador treinado usando esse conjunto de dados aprimorado. Esses protocolos têm como base as métricas de avaliação de modelos considerados estado-da-arte, como acurácia, AUC e F1, por exemplo, comparados a performance do modelo de Inteligência Artificial usando apenas dados reais, ou seja, treinado e testado em conjunto de dados observacionais. Dependendo da forma como os dados sintéticos impactam o modelo de Inteligência Artificial, podemos inferir se esses dados funcionam como complemento do conjunto de treinamento ou como ruído. Utilizaremos três critérios para avaliar a qualidade dos dados sintéticos: Fidelidade (precisão), Diversidade e Generalização (SINGH; KUMAR; KHANNA; LAIRD *et al.*, 2024). Fidelidade é definida como a medida da qualidade das amostras sintéticas em relação a um modelo treinado. Diversidade refere-se à extensão em que essas amostras abrangem a variabilidade total das amostras reais. Por outro lado, a Generalização quantifica até que ponto um modelo treinado é capaz de se ajustar aos dados de treinamento sem super ajustar ou copiar esses dados (overfitting) (SINGH; KUMAR; KHANNA; LAIRD *et al.*, 2024).

Para avaliar esses critérios, são empregados testes estatísticos como o teste de Kolmogorov-Smirnov, que verifica se os conjuntos de dados sintéticos seguem a mesma distribuição que os dados reais. Além disso, serão utilizadas métricas estatísticas como média e desvio padrão para avaliar a generalização dos dados sintéticos em relação aos dados reais

Banco de dados

Os dados gerados no estudo serão tratados e anonimizados, de acordo com a Resolução CNS 466/12, Resolução CNS 738/2024 e Lei Geral de Proteção de Dados (Lei nº 13.709/2018), sendo definido que o controlador do banco de dados será o pesquisador principal. Todos os dados coletados para este

estudo ficarão em ambiente controlado pela UFPE por pelo menos 05 (cinco) anos após a conclusão desse estudo.

Resultados esperados

- Espera-se contribuir para a Ciência e Tecnologia (C&T) com a criação de tripletes de forma a serem utilizados como base para ajustes em dispositivos já existentes e assegurar que os modelos de Inteligência Artificial desenvolvidos para o diagnóstico pré-clínico de doenças cardiovasculares sejam treinados com conjuntos de dados de alta qualidade, resultando em ferramentas mais robustas, precisas e eficazes para a detecção precoce dessas condições. Desta forma, nossos resultados irão fornecer ferramentas tecnológicas de inteligência artificial para o diagnóstico precoce de doenças cardiovasculares, levando o estado de Pernambuco a ser um modelo para todo o Brasil.

- Em termos de internacionalização, este projeto irá possibilitar a consolidação da parceria já existente com o Laboratório CARMeN da Université de Lyon – França através do intercâmbio de pesquisadores e estudantes. Da mesma forma, pretende-se testar este modelo em outras populações, como por exemplo, Moçambique e Cabo Verde.

- Em termos de extensão, prevemos visitas aos municípios da Zona da Mata Sul de Pernambuco para oferecer **Eventos de Ciência Aberta** com oficinas, workshops, palestras e orientações em relação a importância de um estilo de vida saudável através do consumo de alimentos saudáveis e a prática de atividade física regular de forma a combater o aparecimento precoce de doenças cardiovasculares.

- Como plano para popularização científica e tecnológica, pretendemos estabelecer parcerias com empresas e com a secretaria de saúde de forma a oferecer novos dispositivos *wearables* para melhorar o monitoramento dos indicadores de saúde do Estado de Pernambuco.

- Por fim, espera-se a publicação de 5 artigos científicos em revistas de alto impacto indexada internacional na área de ciência da computação aplicada à saúde com parceria com outras instituições de pesquisa nacionais e internacionais.

Aderência aos Objetivos do Desenvolvimento Sustentável (ODS)

Este projeto se enquadra no **ODS #3 Saúde e Bem-Estar**. Neste ODS, estamos na meta 3.4 – Redução da mortalidade por doenças não-comunicáveis e promover saúde mental. Em 2030, pretende-se reduzir em um terço a taxa de mortalidade por doenças não-comunicáveis através da prevenção e tratamento.

3 Orçamento

Quadro de desembolso

Item Orçado	Rubrica			Justificativa	Valor (R\$)
	Capital	Custeio	Bolsa		
Eletrocardiograma Equivital Wearable ECG Starter Pack	X			Avaliação da variabilidade cardíaca. Equipamento de alta precisão que irá subsidiar a validação dos dados do cluster fisiológico	R\$ 50.000,00
LabChart 8 Pro Software (for Windows and Mac).	X			Software utilizado para o Eletrocardiograma Equivital	R\$ 59.300,00
Equipamento Bioimpedanciômetro profissional Marca: InBody - Modelo: InBody270	X			Avaliação de todas as medidas antropométricas e de composição corporal. Este modelo é portátil e de última geração.	R\$ 44.800,00
Material de rotina de laboratório		X		Material de rotina para as coletas: luvas, máscaras, pró-pé, kit para dosar glicemia, tubos, fitas reagentes e caixas de descarte	R\$ 13.200,00
Notebook 13ª geração Intel® Core™ vPro® Enterprise i9-13900H (14-cores, 20 threads, cache de 24MB, 2.6GHz até 5.4GHz, 45W)	X			Computador portátil para a coleta de dados fora do Laboratório de forma a rodar os programas de computação científica.	R\$ 13.380,00
Sistema LDX analyser - Cholestech (colesterol e frações, triglicerídeos e proteínas)	X			Equipamento portátil utilizado para dosar colesterol e frações, triglicerídeos e proteínas. Será utilizado para validação do cluster de indicadores bioquímicos	R\$ 9.200,00
Fitas reagentes para o Cholestech LDX		X		Fitas reagentes para utilizar no cholestech LDX	R\$ 5.100,00
MAPA monitor pressão arterial contec abpm50 com software	X			Monitor de pressão arterial ambulatorial para ser utilizado por cada paciente durante 24h. Será utilizado para validar os dados de pressão arterial sistólica e diastólica. Serão adquiridos 5 monitores.	R\$ 8.000,00
Despesas Acessórias		X		Taxa de instalação dos equipamentos e treinamento da equipe para o uso o Eletrocardiograma e para balança de bioimpedância.	R\$ 5.380,00
Passagens aéreas		X		3 visitas (a cada 6 meses) de 3 pesquisadores ligados ao projeto ao Laboratório Nacional de Computação Científica (LNCC) no Rio de Janeiro.	R\$ 25.000,00
Diárias		X		Cada visita (total de 3) dos pesquisadores ao LNCC terá duração de 5 dias.	R\$ 14.400,00
Bolsa BCT-3			X	Bolsista de cooperação técnica durante 10 meses para compor a equipe de pesquisa na coleta e análise de dados e produção científica.	R\$ 52.000,00
Total					R\$ 299.760,00

3.2 Cronograma físico-financeiro

Setembro de 2024 a janeiro de 2025 = Aquisição da balança de bioimpedância e do Cholestec LDX e fitas reagentes e compra de parte do material de rotina de laboratório. Total de desembolso = R\$ 66.580,00

Setembro de 2024 a Julho de 2025 = Quota do bolsista BCT-3 por 10 meses. Total de desembolso = R\$ 52.000,00

Outubro de 2024 a Junho de 2025 = Coleta de dados do estudo 1. Aquisição do notebook para a coleta dos dados. Total de desembolso = R\$ 13.380,00

Mai de 2025 = Aquisição do Eletrocardiograma Equivital, software LabChart e suas respectivas despesas acessórias. Total de desembolso = R\$ 119.300,00

Julho a dezembro de 2025 = Coleta de dados do estudo 2. Compra dos monitores MAPA e o restante do material de rotina de laboratório. Total de desembolso = R\$ 9.100,00

Janeiro a dezembro de 2026 = Visitas ao Laboratório Nacional de Computação Científica para elaboração dos clusters. Total de desembolso = R\$ 39.400,00

Abril de 2026 = parte da equipe irá visitar os municípios da zona da mata sul de Pernambuco para as atividades de extensão e popularização científica e tecnológica. Será elaborada uma cartilha para ampla divulgação sobre promoção da saúde e bem-estar.

Março a Julho de 2026 = Análise Topológica de Dados (TDA), Construção de bases sintéticas dos dados.

Agosto de 2026 = Apresentação e entrega dos modelos de IA para o diagnóstico pré-clínico de hipertensão arterial e diabetes mellitus tipo 2 para a comunidade científica e para a Secretaria de Saúde do Estado.

Agosto e Setembro de 2026 = Publicação dos resultados em revistas científicas internacionais de alto impacto.

3.3 Contrapartida

O Centro Acadêmico de Vitória (CAV-UFPE) tem infraestrutura para a validação dos dados coletados no HC/EBSERH como o Laboratório de Captura de Sinais Biológicos equipado com Sistema de Densitometria Óssea por Raio-X (DEXA) que é o padrão-ouro para avaliação de composição corporal. Também temos o Laboratório Multiusuário de Nutrição, Atividade Física e Plasticidade Fenotípica equipado com centrífugas, balanças de precisão, freezer para armazenar amostras, etc. Para as coletas de dados, o HC/EBSERH disponibilizou o setor de cardiologia e endocrinologia para instalação dos equipamentos portáteis que serão adquiridos. Por fim, o Centro de Informática (Cin) da UFPE irá disponibilizar os equipamentos necessários para análise dos dados em computadores científicos.

4 Equipe

Carol Virgínia Góis Leandro, coordenadora do projeto. É Professora Titular do Centro Acadêmico de Vitória (curso de Nutrição) da UFPE (CAV-UFPE). É bolsista de produtividade em pesquisa nível 1D pelo comitê de Nutrição e Saúde Coletiva do CNPq. Membro da Academia Pernambucana de Ciências. Atualmente é Pró-reitora de Pós-graduação da UFPE. É membro permanente do PPG em Nutrição, Atividade Física e Plasticidade Fenotípica (CAV-UFPE, nível 4 da CAPES), do PPG Multicêntrico em Ciências Fisiológicas do CAV-UFPE (nota 5 da CAPES), e do PPG em Nutrição da UFPE (campus Recife, nível 5 da CAPES). Coordena o projeto internacional CAPES/COFECUB (Edital 932-20)": Nutrição e atividade física no combate à obesidade: da clínica à biologia molecular, projeto "VITORIA" entre a UFPE e a Université de Lyon – França.

Atuação no projeto = coordenadora geral, aquisição de equipamentos e reagentes. Coleta e análise dos dados em cada cluster.

Cleber Zanchettin, membro da equipe executiva. É Professor Associado do Centro de Informática da UFPE (Cin-UFPE). Bolsista de Produtividade Desen. Tec. e Extensão Inovadora do CNPq - Nível 2. É Professor Visitante na Northwestern University, EUA. Membro permanente do PPG em Ciências da Computação da UFPE (nota 7 da CAPES). Sua linha de pesquisa inclui *Deep Learning*, Redes Neurais Artificiais e Segurança Cibernética. É mentor do Lócus de Inovação em Inteligência Artificial e Cibersegurança do Estado de Pernambuco. É também um membro ativo de comitês como o da AMCHAM para Inteligência Artificial e o de regulamentação da IA no Brasil pela 3 Câmara de Coordenação e Revisão do MPF. Colabora em projetos com empresas de renome como Samsung, Tempest e Embraer.

Atuação no projeto = Aplicação dos protocolos de ampliação de dados reais, análise dos dados sintéticos, irá realizar os experimentos com inteligência artificial, geometria e análise topológica dos dados.

Adriano Augusto de Moraes Sarmento. Doutor em Nano et Micro electronique - Universite de Grenoble I (Scientifique Et Medicale - Joseph Fourier) (2005). Atualmente é professor associado do Centro de Informática da Universidade Federal de Pernambuco, sendo integrante do grupo de pesquisa GRECO (Grupo de Engenharia da Computação). Membro permanente do PPG em Ciências da Computação da UFPE (nota 7 da CAPES). Tem experiência em Arquitetura de Sistemas de Computação, ambientes de projeto de sistemas embarcados, visão computacional embarcada, internet das coisas, modelagem, síntese de interfaces hardware/software, co-simulação de sistemas heterogêneos.

Atuação no projeto = Aplicação dos protocolos de inteligência computacional, anonimização dos dados e modelagem.

Wylla Tatiana Ferreira e Silva, membro da equipe executiva. É Professora Associada do Núcleo de Nutrição do CAV-UFPE. Nutricionista, Mestre e Doutora em Nutrição. Docente permanente do PPG em Nutrição, Atividade Física e Plasticidade Fenotípica do CAV-UFPE. Atua na área de: alimentação saudável, segurança alimentar, impactos da alimentação ultraprocessada na saúde humana; alimentação e sustentabilidade; direito humano à alimentação.

Atuação no projeto = Avaliação do consumo alimentar, frequência alimentar e irá liderar as visitas aos municípios, as atividades de extensão e popularização científica e tecnológica.

Declaração

Eu, Carol Virgínia Góis Leandro, declaro para os devidos fins de direito que não possuo qualquer inadiplência com a FACEPE e com a Administração Pública Federal, direta ou indireta.

Referências

AL ASHMAR, S.; ANWARDEEN, N. R.; ANLAR, G. G.; PEDERSEN, S. *et al.* Metabolomic profiling reveals key metabolites associated with hypertension progression. **Front Cardiovasc Med**, 11, p. 1284114, 2024.

CHARCHAR, F. J.; PRESTES, P. R.; MILLS, C.; CHING, S. M. *et al.* Lifestyle management of hypertension: International Society of Hypertension position paper endorsed by the World Hypertension League and European Society of Hypertension. **J Hypertens**, 42, n. 1, p. 23-49, Jan 1 2024.

ENDRIAS, E. E.; TESFAYE MAMITO, T.; GETA HARDIDO, T.; ATARO, B. A. Prevalence of Pre-Hypertension/Hypertension and Its Associated Factors Among Adults in the Wolaita Zone of Southern Ethiopia: A Cross-Sectional Study. **Inquiry**, 61, p. 469580241246968, Jan-Dec 2024.

FEITOSA, A. D. M.; BARROSO, W. K. S.; MION JUNIOR, D.; NOBRE, F. *et al.* Brazilian Guidelines for In-office and Out-of-office Blood Pressure Measurement - 2023. **Arq Bras Cardiol**, 121, n. 4, p. e20240113, Feb 2024.

JULIAO, N. A.; SOUZA, A.; GUIMARAES, R. R. M. Trends in the prevalence of systemic arterial hypertension and health care service use in Brazil over a decade (2008-2019). **Cien Saude Colet**, 26, n. 9, p. 4007-4019, Sep 2021.

KANAZAWA, I.; SUGIMOTO, T. Prehypertension increases the risk of atherosclerosis in drug-naive Japanese patients with type 2 diabetes mellitus. **PLoS One**, 13, n. 7, p. e0201055, 2018.

MARTINEZ-MURCIA, F. J.; ARCO, J. E.; JIMENEZ-MESA, C.; SEGOVIA, F. *et al.* Bridging Imaging and Clinical Scores in Parkinson's Progression via Multimodal Self-Supervised Deep Learning. **Int J Neural Syst**, 34, n. 8, p. 2450043, Aug 2024.

MAYORAL, L. P.; ANDRADE, G. M.; MAYORAL, E. P.; HUERTA, T. H. *et al.* Obesity subtypes, related biomarkers & heterogeneity. **Indian J Med Res**, 151, n. 1, p. 11-21, Jan 2020.

MONIKA; ARORA, P.; KUMAR, V.; POPLI, P. Conceptual Overview of Prevalence of Prediabetes. **Curr Diabetes Rev**, Apr 3 2024.

MUZY, J.; CAMPOS, M. R.; EMMERICK, I.; SILVA, R. S. D. *et al.* [Prevalence of diabetes mellitus and its complications and characterization of healthcare gaps based on triangulation of studies]. **Cad Saude Publica**, 37, n. 5, p. e00076120, 2021.

NAKANGA, W. P.; CRAMPIN, A. C.; MKANDAWIRE, J.; BANDA, L. *et al.* Waist circumference and glycaemia are strong predictors of progression to diabetes in individuals with prediabetes in sub-Saharan Africa: 4-year prospective cohort study in Malawi. **PLOS Glob Public Health**, 3, n. 9, p. e0001263, 2023.

NANDAGOPAL, M.; SEERANGAN, K.; GOVINDARAJU, T.; ABI, N. E. *et al.* A Deep Auto-Optimized Collaborative Learning (DACL) model for disease prognosis using AI-IoMT systems. **Sci Rep**, 14, n. 1, p. 10280, May 4 2024.

OBRYCKI, L.; SARNECKI, J.; PAC, M.; DEREZINSKI, T. *et al.* Accelerated vascular age in adolescents with primary hypertension. **J Hypertens**, 41, n. 1, p. 171-179, Jan 1 2023.

PARANJPE, I.; WANG, X.; ANANDAKRISHNAN, N.; HAYDAK, J. C. *et al.* Deep learning on electronic medical records identifies distinct subphenotypes of diabetic kidney disease driven by genetic variations in the Rho pathway. **medRxiv**, Sep 7 2023.

RAO, L.; LU, J.; WU, H. R.; ZHAO, S. *et al.* Automatic classification of fetal heart rate based on a multi-scale LSTM network. **Front Physiol**, 15, p. 1398735, 2024.

SINGH, M.; KUMAR, A.; KHANNA, N. N.; LAIRD, J. R. *et al.* Artificial intelligence for cardiovascular disease risk assessment in personalised framework: a scoping review. **EClinicalMedicine**, 73, p. 102660, Jul 2024.

SRIVASTAVA, A.; MIRZA, T. M.; VAQAR, S.; SHARAN, S. Prehypertension. *In*: **StatPearls**. Treasure Island (FL), 2024.

TROTTET, C.; ALLAM, A.; HORVATH, A. N.; FINCKH, A. *et al.* Explainable deep learning for disease activity prediction in chronic inflammatory joint diseases. **PLOS Digit Health**, 3, n. 6, p. e0000422, Jun 2024.

VYAS, A.; DESHPANDE, A.; SEN, S.; KIM, R. *et al.* Population-Based Artificial Intelligence Assessment of Relationship Between the Risk Factors for Diabetic Retinopathy in Indian Population. **Ophthalmic Epidemiol**, p. 1-7, Dec 12 2023.

XUAN, Q.; HU, C.; ZHANG, Y.; WANG, Q. *et al.* Serum lipidomics profiles reveal potential lipid markers for prediabetes and type 2 diabetes in patients from multiple communities. **Front Endocrinol (Lausanne)**, 13, p. 966823, 2022.