

Sistema de Apoio à Prevenção da Recidiva do Câncer de Mama com Algoritmos de Otimização e Segmentação de Imagens

Carlos Estellita Neto, Thalyson Gomes Nepomuceno da Silva, Gustavo Augusto Lima de Campo,

Thelmo Pontes de Araujo, Sara Maria Moreira Lima Verde

Universidade Estadual do Ceará - Fortaleza, Brasil

carlos.estellita@aluno.uece.br, thalyson.uece@gmail.com, gustavo.campos@uece.br,

thelmo.araujo@uece.br, sara.maria@uece.br

Abstract—This study proposes the development of an intelligent system to prevent breast cancer recurrence, combining aspects of nutritional assessment, medical image segmentation and local search optimization. From the absorption of data by the ANPREC questionnaire, together with information from computed tomography segmentation, optimization algorithms, Genetic Algorithm and NSGA-II, personalized plans of recommendations are formulated to improve the patient's healthy habits. In the end, it can be concluded that NSGA-II presents a good formulation of personalized plans; however, for the system to be more effective, it is necessary to develop a more robust segmentation method.

Keywords—Breast Neoplasms; Recurrence; Optimization Algorithms; Image Segmentation.

Resumo—Este estudo propõe o desenvolvimento de um sistema inteligente para prevenção da recidiva do câncer de mama, combinando aspectos de avaliação nutricional, segmentação de imagens médicas e otimização em busca local. A partir da absorção dos dados pelo questionário ANPREC, juntamente com as informações provenientes da segmentação da tomografia computadorizada, algoritmos de otimização, Algoritmo Genético e NSGA-II, formulam planos personalizados de recomendações para melhoria de hábitos saudáveis do paciente. Ao final, pode-se concluir que o NSGA-II apresenta boa formulação de planos personalizados; porém, para o sistema possuir melhor eficácia, é necessário desenvolver um método de segmentação mais robusto.

Palavras-chave—Câncer de Mama; Recidiva; Algoritmos de Otimização; Segmentação de Imagem.

I. INTRODUÇÃO

O câncer de mama é a segunda neoplasia mais recorrente no mundo, com um aumento para 18,1 milhões de novos casos e 9,6 milhões de mortes no ano de 2018. Dentre os tipos oncológicos, o câncer de mama é o segundo mais recorrente no mundo e o primeiro entre mulheres [1]. No Brasil, este cenário é semelhante ao mundial, onde 29,5% das mulheres possuem câncer de mama, sendo o mais comum dentre as neoplasias [2].

Até o fim de 2020, sobreviventes de câncer de mama em todo o mundo chegavam a 7 milhões dentro de 5 anos após o diagnóstico [3]. A recidiva tardia, que pode ocorrer entre 5 e 20 anos após o diagnóstico, é motivo de preocupação. Desta forma, torna-se de extrema necessidade que haja um diagnóstico precoce eficiente, pois quando os casos são detectados e tratados previamente, a mortalidade por câncer é reduzida [4], ajudando a uma sobrevivida maior.

Um dos fatores de relevância para o desenvolvimento de câncer de mama é o elevado IMC ($\frac{kg}{m^2}$), principalmente quando está associado ao aumento na quantidade de gordura [5], [6]. O excesso de gordura conduz ao estado de inflamação crônica de baixa intensidade, com maior quantidade de citocinas inflamatórias circulantes e desbalanço entre as adipocitocinas as quais estimulam as cascatas carcinogênicas de proliferação celular e inibem a morte celular programada [6]. De acordo com [7], o não gerenciamento do peso e alimentação entre sobreviventes pode cancelar os efeitos benéficos do tratamento dessas mulheres e favorecer a recidiva da doença.

Segundo o IBGE, a proporção de mulheres com 20 ou mais anos de idade acima do peso mais que dobrou no Brasil, tendo um aumento de 14,5% para 30,2% entre os anos de 2003 e 2019. Além disso, quanto maior a faixa etária, maior é o predomínio de sobrepeso. Por exemplo, 57,0% das mulheres são vítimas do excesso de gordura na faixa de 25 a 39 anos [8].

Em 2018, o WCRF/IARC apresentou 10 recomendações para prevenção do câncer, que envolvem alimentação e atividade física, como: 1. manter o peso saudável; 2. ser fisicamente ativo; 3. consumir grãos e cereais integrais, frutas, legumes e feijões; 4. limitar o consumo de fast food e alimentos ricos em açúcares e gorduras; 5. limitar o consumo de carne vermelha e processada; 6. limitar o consumo de bebida açucarada; 7. não

usar suplementos para a prevenção do câncer [9]. Ademais, como já comentado anteriormente, um conhecimento mais aprofundado sobre o percentual de gordura do paciente é essencial, pois se trata de um fator relevante para ocorrer a recidiva de câncer.

Nesta perspectiva, surge o uso de sistemas computacionais de processamento da informação com o objetivo de aproximar o usuário da informação correta, torná-lo autônomo, permitir sua auto percepção sobre o estado de saúde e favorecer o autocuidado. Este trabalho apresenta um sistema computacional capaz de gerar planos de recomendações personalizadas para direcionar pacientes à mudança de seus hábitos de saúde e estilo de vida e promover melhor a qualidade de vida, a partir de um diagnóstico individualizado de prevenção da recidiva de câncer de mama, empregando informações sobre fatores modificáveis de peso, alimentação, atividade física e composição corporal.

O sistema foi projetado para o uso no acompanhamento clínico, onde profissionais da saúde e pacientes interagem com um questionário para informar os dados essenciais. A partir dessas informações, o sistema estima o risco e gera planos de prevenção à recidiva do câncer de mama.

Este projeto possui três objetivos específicos para que assim possa ser gerado o sistema inteligente proposto:

- **Segmentação de Imagens:** É de suma importância segmentar os diferentes tipos de gordura (subcutânea e visceral, que estão sendo abordadas neste trabalho) presentes nas imagens de tomografias abdominais, pois como cada gordura possui uma diferente classificação de risco à saúde, identificar precisamente as quantidades e proporções de gorduras é um fator relevante no momento de classificar o estado de saúde do paciente. Este trabalho se propõe a segmentar gordura subcutânea e visceral.
- **Algoritmo de Otimização:** Com base nas informações adquiridas do paciente através do ANPREC, juntamente com as informações das gorduras adquiridas pela imagem de tomografia, o paciente receberá uma classificação de saúde. Com base em sua classificação e no escore de saúde desejado pelo profissional de saúde em conjunto com o paciente, será implementado um algoritmo de otimização com o intuito de gerar um plano de hábitos saudáveis de forma personalizada.
- **Desenvolvimento Web:** Por fim, todas estas aplicações serão apresentadas em conjunto num ambiente virtual *web*, onde o paciente e profissional de saúde poderão responder ao questionário ANPREC, importar a imagem de tomografia, receber sua segmentação, receber uma classificação de saúde e ser apresentado o plano de recomendações gerado.

II. TRABALHOS RELACIONADOS

Nesta seção, são apresentadas pesquisas que contribuem para a área de desenvolvimento de sistemas inteligentes para geração de recomendações. Este tópico possui a finalidade de apresentar a grande variedade de metodologias para geração de recomendações personalizadas no âmbito da saúde.

A seleção de artigos foi realizada através de buscas em portais acadêmicos, como no Portal de Periódicos da CAPES, SpringerLink e procurando em referenciais teóricos em artigos selecionados.

A. A framework of hybrid recommender system for personalized clinical prescription

No artigo desenvolvido em [10], é apresentada uma proposta de como auxiliar médicos clínicos gerais na tomada de decisão durante a prescrição de medicamentos, abordando a sobrecarga de informações devido ao surgimento constante de novos medicamentos e suas complexas interações. Foi desenvolvido um *framework* para identificar as necessidades dos pacientes com base em seus sintomas, extrair características pertinentes de textos e analisar a eficácia dos medicamentos, garantindo prescrições mais precisas e alinhadas às condições do paciente.

A metodologia do artigo foi dividida em duas fases: treinamento e aplicação. Na fase de treinamento, o sistema extrai características de textos médicos usando técnicas de processamento de linguagem natural para identificar sintomas e classificá-los de acordo com sua gravidade. Esses dados são combinados com resultados estruturados de exames laboratoriais, formando um espaço de características do paciente. Paralelamente, medicamentos são agrupados em *clusters* com base nos sintomas, usando o algoritmo de *k-means*. Em seguida, uma rede neural é treinada para prever quais *clusters* de medicamentos são mais adequados a diferentes combinações de sintomas. Durante a fase de aplicação, o sistema recebe novas informações de pacientes e utiliza o modelo treinado para recomendar medicamentos, priorizando aqueles com maior similaridade a casos anteriores, além de considerar o impacto temporal do uso do medicamento.

Os resultados do *framework* indicam melhorias significativas na precisão das recomendações e na eficiência do processo de prescrição, minimizando erros e reduzindo o tempo necessário para que os médicos possam tomar decisões assertivas. Os autores concluem que o sistema proposto preenche lacunas existentes em pesquisas anteriores ao considerar o relacionamento complexo entre sintomas e medicamentos.

B. The Runner - Recommender System of Workout and Nutrition for Runners

Segundo os autores em [11], é apresentado um sistema de recomendação de treino e alimentação para corredores

profissionais e amadores, capaz de integrar o conhecimento de especialistas com a experiência social de outros corredores, utilizando técnicas da *web* semântica e ontologias para formular soluções completas e personalizadas.

Os autores desenvolveram uma metodologia que envolve a construção de um sistema de três camadas: interface de usuário, lógica de negócios e banco de dados. A interface foi desenvolvida em Java Server Faces, permitindo que o usuário crie um perfil e defina metas. A camada de lógica de negócios, implementada em Java Beans, é responsável por processar as consultas e adaptar planos de treino e dietas conforme o desempenho do usuário. A base de dados, armazenada em MySQL, utiliza ontologias OWL que organizam o conhecimento em três domínios: nutrição, medicina esportiva e corrida. A construção das ontologias contou com a colaboração de especialistas e foi populada por meio de *web crawling* utilizando o *framework* Apache Nutch para fazer a extração de informações de sites relevantes. A adaptação do sistema é contínua, recalculando planos de treino em caso de lesões ou falhas nos treinos, reforçando a flexibilidade do modelo.

Ao final, é possível observar que a combinação entre *web* semântica, ontologias e recomendação social proporciona um sistema robusto e personalizado, capaz de atender desde corredores iniciantes até atletas intermediários. Testes com usuários indicaram que a personalização das recomendações e a simplicidade na interface foram pontos de destaque, garantindo adesão e engajamento. A principal contribuição do sistema está na capacidade de adaptar não apenas planos de treino, mas também dietas específicas para cada fase do treinamento, promovendo um ciclo completo de preparação física e recuperação.

C. PRO-Fit: A personalized fitness assistant framework

Os autores em [12] desenvolveram um *framework* que utiliza tecnologias vestíveis, como acelerômetros e GPS, e algoritmos de aprendizado de máquina para oferecer recomendações personalizadas de atividades físicas. O sistema visa rastrear a atividade do usuário e sincronizar essas informações com o calendário pessoal para sugerir sessões de treino adaptadas ao estilo de vida, localização e disponibilidade de cada usuário.

A metodologia do sistema é composta por um aplicativo móvel que coleta os dados dos sensores do dispositivo do usuário, sincronizando com calendários digitais para organizar sessões de treino. O aplicativo é responsável pela comunicação com servidores externos por meio de Interfaces de Programação de Aplicativos, APIs, enquanto a camada de pré-processamento classifica os dados coletados em janelas temporais, utilizando técnicas de streaming de dados (com Apache Kafka) e armazenamento em Cassandra. A classificação das atividades é realizada através de algoritmos como Decision Tree e Random Forest, que identificam atividades específicas como corrida,

caminhada e ciclismo. Por fim, o sistema consegue sugerir sessões de treino personalizadas.

Ao final do artigo, os pesquisadores concluem que o *framework* oferece uma boa abordagem para o incentivo à prática de atividades físicas, automatizando o processo de planejamento de treinos e reduzindo a necessidade de interação constante por parte do usuário.

D. A Recipe Recommendation System based on Automatic Nutrition Information Extraction

No estudo presente em [13] os pesquisadores apresentam um sistema de recomendação de receitas, cujo diferencial está na extração automática de informações nutricionais a partir de dados disponíveis na internet. A pesquisa visa oferecer aos usuários uma ferramenta que possibilita a busca de receitas que atendem a necessidades de saúde específicas, utilizando processamento de linguagem natural. Diferente de sites convencionais que exigem conhecimento prévio sobre nutrição, a abordagem proposta democratiza o acesso a dietas mais equilibradas, permitindo que qualquer usuário encontre receitas que possam auxiliar na prevenção ou no tratamento de condições como fadiga, acne ou diabetes.

O sistema é composto por cinco etapas principais que integram técnicas de processamento de linguagem natural e análise nutricional automatizada. Primeiramente, o usuário insere um problema de saúde em linguagem simples, como “Quero melhorar minha pele”. O sistema, então, realiza uma análise morfológica da frase, identificando palavras-chave relacionadas a condições de saúde. Essas palavras são pesquisadas em um banco de dados de coocorrência, que correlaciona 45 nutrientes com termos comuns ligados à saúde. A partir dessa busca, o sistema identifica o nutriente mais relevante para a condição indicada e, em seguida, pesquisa no banco de dados de receitas por pratos que contenham ingredientes ricos nesse nutriente. A base de receitas, composta por 800.000 pratos extraídos do site Cookpad, permite uma ampla gama de resultados, os quais são exibidos ao usuário de forma intuitiva em dispositivos móveis.

A pesquisa revelou que usuários com pouco conhecimento em nutrição encontraram valor na ferramenta, embora houvesse uma percepção de que pratos com vegetais são automaticamente saudáveis.

E. Yum-me: A Personalized Nutrient-based Meal Recommender System

O estudo proposto em [14] desenvolve um sistema de recomendação de refeições personalizadas, com base nas preferências alimentares dos usuários e nas suas necessidades nutricionais, a fim de controlar doenças como diabetes e obesidade. O sistema permite que os usuários expressem suas preferências através de uma interface visual intuitiva, sem a

necessidade de registros detalhados de refeições passadas, podendo até mesmo citar restrições alimentares, como veganismo e kosher, por exemplo.

O usuário responde a um questionário indicando restrições dietéticas e metas nutricionais. Essa etapa inicial gera um conjunto de pratos candidatos extraídos de um banco de dados com cerca de 50.000 receitas. Em seguida, é feito um quiz visual, no qual o usuário escolhe pratos que parecem mais atraentes, permitindo ao sistema refinar as recomendações de acordo com preferências mais detalhadas. Para isso, é utilizado o modelo FoodDist, uma rede neural convolucional que analisa imagens de alimentos e mede a similaridade entre pratos.

Os resultados obtidos demonstram a eficácia do sistema em gerar recomendações alinhadas ao gosto e às necessidades nutricionais do usuário. Em um estudo com 60 participantes, o Yum-me aumentou em 42,63% a taxa de aceitação das recomendações em comparação a métodos tradicionais baseados em questionários. O artigo conclui que o Yum-me oferece uma alternativa eficaz e interativa para promover hábitos alimentares mais saudáveis.

F. Panorama

A intenção no tópico de Trabalhos Relacionados está em demonstrar a variabilidade de metodologias de geração de planos para condicionar melhor qualidade de vida aos pacientes. O pioneirismo do presente trabalho está em utilizar métodos de otimização de busca local juntamente com processamento de imagens, para formular planos mais bem adaptados aos pacientes, contrastando com outros trabalhos da literatura que não realizam uma abordagem tão plural no âmbito de inteligência computacional.

III. METODOLOGIA

Neste presente trabalho, é proposto um sistema capaz de fazer o diagnóstico individualizado de pacientes com base em características mutáveis, relacionadas a peso, práticas de exercícios físicos, alimentação e composição corporal, e gerar planos de recomendações personalizados para direcionar esses pacientes à mudança em seus hábitos de saúde, estilo de vida e promover um bem-estar mais saudável.

Esta seção apresenta como foi englobada a segmentação de imagens, os algoritmos de otimização e o questionário ANPREC em uma plataforma *web* de auxílio aos profissionais de saúde e pacientes. Abaixo segue a imagem que exemplifica a arquitetura do sistema e como cada aspecto está relacionado.

O paciente e o médico, através de uma API de comunicação, respondem ao formulário ANPREC oferecendo dados do paciente. Além das respostas do ANPREC, o sistema oferece a opção de ser importada uma imagem para ser feita a segmentação. Com base nas informações do ANPREC e

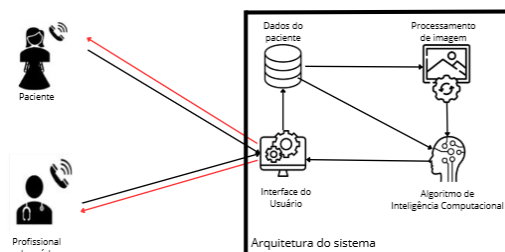


Fig. 1. Diagrama da arquitetura do sistema elaborada pelo autor.

da segmentação da imagem de tomografia, o algoritmo de otimização formula um plano de recomendações e apresenta aos usuários, paciente e profissional da saúde, através da interface do usuário.

A. ANPREC

O ANPREC é dividido em oito seções de perguntas, relacionadas ao estado atual do paciente ou às suas ações:

- Questionamento sobre o estado de peso atual do paciente:
 - Aumento de peso na idade adulta.
 - IMC atual.
 - Circunferência da cintura.
- Práticas de atividade física:
 - Prática de atividade física diária.
 - Duração e intensidade da sua atividade física diária.
 - Gasto de mais de três horas por dia com hábitos sedentários.
- Práticas de alimentação:
 - Consumo de chocolates, sorvetes, industrializados, bolos com recheio e afins.
 - Consumo de refrigerante e/ou outras bebidas açucaradas.
 - Consumo de frituras, salgadinhos e afins.
- Práticas de alimentação de origem vegetal:
 - Consumo de frutas de cores variadas, verduras e hortaliças sem amido.
 - Cereal integral e /ou leguminosa.
- Práticas de alimentação de origem animal:
 - Porções de carne vermelha.
 - Consumo de carne processada.
- Práticas de consumo de álcool:
 - Porções de bebida alcoólica.
- Práticas de consumo de alimentos processados e preparados com adição de sal:
 - Uso de temperos prontos.
 - Consumo de produtos industrializados.

- Costume de adicionar sal em refeições.
- Práticas de consumo de suplementos:
 - Consumo de suplemento alimentar.

Cada questão gera uma pontuação, que com base na resposta do paciente, pode ser **0**, **0,5** ou **1** escore. Esses pontos são somados para calcular o escore total. A classificação de risco possui a seguinte escala:

- **Escore ≤ 9 :** Risco aumentado de recidiva
- **$9 < \text{Escore} < 12$:** Risco moderado
- **Escore ≥ 12 :** Risco reduzido

É importante ressaltar que algumas perguntas, relacionadas ao peso e medidas antropométricas, são excluídas do processo de recomendação, sendo usadas apenas para avaliação de risco. Isso ocorre porque essas perguntas tratam do estado do paciente e não refletem em uma ação que pode ser recomendada. As demais perguntas sobre os hábitos do paciente, no entanto, são utilizadas pelos algoritmos de otimização para formular os planos.

B. Algoritmos de otimização

Com base na pontuação feita em cada pergunta, classificação de saúde do paciente e classificação de saúde desejada, será formulado um *array* onde cada elemento será a respectiva pontuação do questionário. O *array* atua como cromossomo e seus elementos como genes. Dessa forma, tanto o AG quanto o NSGA-II são implementados. O escore atual define o ponto de partida para a geração de recomendações personalizadas e o profissional da saúde escolhe um escore para ser atingido através do plano.

1) *Uso do Algoritmo Genético no ANPREC:* Com o cromossomo formulado, o AG simula a evolução através de seleção, *crossover* e mutação, com o objetivo de minimizar a diferença entre o escore atual e o escore desejado.

Inicia-se a geração de uma população inicial composta por indivíduos representados por *arrays* que contêm as respostas do questionário. Cada indivíduo é uma combinação de escores que refletem as respostas do questionário, contendo os valores possíveis de 0, 0,5 ou 1.

O *fitness* é calculado com base na diferença entre o escore do indivíduo e o escore desejado, e indivíduos com menor diferença possuem maior chance de serem selecionados. A seleção dos indivíduos para a próxima geração é realizada por meio de um método de roleta, onde indivíduos com melhor *fitness* têm maior probabilidade de serem escolhidos.

Durante o processo evolutivo, são aplicadas operações de *crossover*, com probabilidade de 50%, permitindo que dois indivíduos troquem partes de seus *arrays*, gerando novas soluções. Além disso, a mutação ocorre com taxa de 20%,

modificando aleatoriamente elementos dos indivíduos para promover diversidade genética.

Esse processo é iterado por 50 gerações para gerar indivíduos cada vez mais próximos do escore desejado ou até ser encontrado um indivíduo que supre o escore requisitado. Este indivíduo final é selecionado como o plano de recomendação final, apresentando as mudanças necessárias nos hábitos do paciente para atingir um maior grau de prevenção da recidiva do câncer.

2) *Uso do NSGA-II no ANPREC:* Assim como no AG, o algoritmo evolui uma população de planos de recomendações através de seleção, *crossover* e mutação, buscando soluções que atinjam o escore desejado e com o menor número de modificações.

O NSGA-II foi implementado de duas formas distintas: uma para minimizar o número de recomendações e outra para minimizar o número de saltos do plano de recomendações.

A escolha de serem implementadas a minimização de números de recomendação e minimização do número de saltos se dá pelo fato de tentar aproximar os planos à realidade do paciente, por exemplo: às vezes um indivíduo não consegue empregar vários hábitos saudáveis cotidianamente, logo é mais plausível que a pessoa escolha um plano que seja possível adquirir o escore desejado sem se sobrecarregar de atividades para fazer durante o dia; ou quando um indivíduo não consegue fazer mudanças muito súbitas em seu estilo de vida, logo é aconselhável ter um plano de recomendações que minimize o número de saltos em seus hábitos.

Na primeira implementação, de minimizar o número de recomendações alteradas, o NSGA-II foi feito para encontrar um equilíbrio entre dois objetivos conflitantes: minimização da diferença entre o escore atual e o escore desejado; e minimização do número de recomendações alteradas.

Inicia-se a geração de uma população inicial composta por indivíduos representados por *arrays* que contêm as respostas do questionário. Indivíduos são gerados aleatoriamente, respeitando as opções disponíveis para cada pergunta.

O *fitness* avalia dois objetivos: a diferença entre o escore do indivíduo e o escore desejado, e o número de respostas alteradas. Indivíduos que atingem o escore desejado com menos mudanças são priorizados.

A seleção é baseada na classificação de Pareto, com indivíduos não-dominados sendo preferidos. A diversidade da população é mantida através de operadores de *crossover*, com probabilidade de 70%, e mutação, com taxa de 20%.

Se um indivíduo não atinge o escore desejado após a seleção, ele é ajustado para maximizar sua pontuação, aumentando gradualmente as respostas que pontuam 0 ou 0,5 até atingir o valor alvo.

O processo é iterado por 50 gerações. Ao final, o indivíduo com a menor diferença entre o escore final e o escore desejado é selecionado como o plano de recomendação otimizado.

Na segunda implementação, de minimizar o número de saltos, o NSGA-II foi feito para encontrar um equilíbrio entre dois objetivos conflitantes: minimização da diferença entre o escore atual e o escore desejado; minimização de saltos bruscos nas recomendações; e minimização do número de recomendações.

É iniciada a geração inicial da população, que é composta por indivíduos representados por *arrays* que contêm as respostas do questionário. Indivíduos são gerados aleatoriamente, respeitando as opções disponíveis para cada pergunta.

O *fitness* avalia cada indivíduo com base nos três objetivos mencionados. Para isso, calcula-se o escore total do indivíduo, a diferença para o escore desejado, o número de recomendações feitas e a quantidade de saltos bruscos.

A seleção é baseada na classificação de Pareto, com indivíduos não-dominados sendo preferidos. A diversidade da população é mantida através de operadores de *crossover*, com probabilidade de 70%, e mutação, com taxa de 20%, alterando a resposta de uma questão para uma opção próxima (por exemplo, de 0 para 0,5 ou de 0,5 para 1), evitando mudanças abruptas.

O processo é iterado por 50 gerações. Ao final, o indivíduo com a menor diferença entre o escore final e o escore desejado é selecionado como o plano de recomendação otimizado.

C. Segmentação de imagens

Há a necessidade de ser feita a segmentação das imagens devido ao fato da pertinência de identificar a proporção entre gorduras para serem utilizadas no processo de classificação da saúde do paciente.

Dessa forma, as imagens que compõem o *dataset* se tratam de tomografias computadorizadas axiais de abdômen na altura da L3 de tipo DICOM (Digital Imaging and Communications in Medicine), totalizando 203 imagens mais as suas respectivas máscaras binárias. O Índice de Massa Corporal, apesar de ser comum para estimar as proporções de massa muscular e de gordura, não apresenta uma correlação significativa com a taxa de sobrevivência de pacientes com câncer. Entretanto, é possível obter correlações melhores com base em proporções das diferentes gorduras. Com isso, foi feita a análise de cada imagem de TC para estimar os percentuais de gordura subcutânea e visceral.

Primeiramente, foi aplicada uma técnica para o aumento de dados, que é um procedimento comum para aumentar a diversidade e a quantidade dos dados de treinamento. Dessa forma, foi feita a expansão do conjunto de dados disponíveis pela implementação em Python nas imagens de TC e suas respectivas máscaras, utilizando a biblioteca *TensorFlow* e a

função *ImageDataGenerator*, que é a responsável por realizar as variações aleatórias das imagens.

A configuração do *ImageDataGenerator* consta dos seguintes fatores:

- **Rotação:** Rotação de até 4 graus.
- **Deslocamento horizontal:** Deslocamento horizontal de até 5% do tamanho da imagem.
- **Deslocamento vertical:** Deslocamento vertical de até 5% do tamanho da imagem.
- **Zoom:** Escala de zoom entre 95% e 105%.
- **Espelhamento:** Espelhamento horizontal.

As transformações são aplicadas de forma consistente às imagens e suas máscaras. Para cada imagem, 17 variações aumentadas são geradas, garantindo que as máscaras sofram as mesmas transformações que as imagens originais, totalizando 3451 imagens únicas por *dataset*.

Para ser feita a segmentação, foi proposta uma abordagem que utiliza a arquitetura de uma Rede Neural Convolutiva, U-Net. A arquitetura da U-Net consiste em uma parte de contração, seguida por uma parte de expansão. A parte de contração da U-Net consiste em repetidos blocos de camadas. Cada bloco na parte de contração possui:

- Uma camada convolutiva com um número de 64 filtros, uma janela de convolução 3x3 com ativação *ReLU* e preenchimento “*same*”, que significa que a saída tem a mesma altura e largura que a entrada.
- A inicialização do *kernel* é feita com o parâmetro “*he_normal*”, que é uma maneira padrão, mas eficiente de inicializar os pesos da rede.
- Camada de *dropout*, com a taxa aumentando progressivamente para camadas mais profundas, iniciando em 0,1 e indo até 0,3.
- Outra camada convolutiva com as mesmas características da primeira.
- Camada de *pooling* máximo para reduzir as dimensões espaciais, aprimorando a invariância à localização e escala da imagem.
- A parte de contração termina com uma camada convolutiva sem o *pooling* máximo, efetivamente formando o fundo do “U”.

Na parte de expansão, cada bloco possui:

- Camada de transposição convolutiva para aumentar as dimensões espaciais da imagem.
- Operação de concatenação que combina a saída da camada de transposição convolutiva com a saída correspondente da camada convolutiva na parte de contração.
- Duas camadas convolutivas, com as mesmas características das da parte de contração.

No final da rede, há uma camada convolucional que gera a saída, com um número de filtros correspondente à quantidade de classes que o modelo precisa identificar. Em seguida, é aplicada uma função *softmax*, que converte esses resultados em probabilidades, indicando a chance de cada pixel pertencer a uma determinada classe.

Os dados de entrada utilizados para o treinamento do modelo consistem em imagens redimensionadas para 256×256 pixels, enquanto as máscaras associadas possuem o mesmo tamanho, representando as áreas de interesse para segmentação. Durante o processo de treinamento, a rede aprende a identificar padrões e características das imagens de entrada, associando-os às máscaras correspondentes, que destacam as regiões pertinentes.

Para treinar o modelo, o conjunto de dados foi dividido em 90% para treino e 10% para teste, garantindo que a rede seja capaz de generalizar bem para novas imagens. O modelo foi otimizado utilizando o algoritmo *Adam*, com uma taxa de aprendizado de 10^{-3} . A função de perda escolhida foi a *binary_crossentropy*, uma vez que o objetivo da U-Net é realizar uma segmentação binária, distinguindo entre a área de interesse e o fundo.

O treinamento foi realizado por 60 épocas, permitindo que o modelo se ajustasse progressivamente aos dados, refinando suas previsões a cada ciclo. A cada época, o desempenho do modelo foi avaliado no conjunto de teste para assegurar que ele estivesse aprendendo de forma consistente e generalizando adequadamente para novos dados.

Após o treinamento, o modelo final foi salvo para ser utilizado na segmentação de novas imagens, com base no aprendizado adquirido durante o processo de treinamento.

O processo de segmentação é utilizado ao final do questionário. Haverá a opção de adicionar uma imagem de tomografia axial de abdômen para ser feita a segmentação. Com base na segmentação, o sistema identificará as proporções de gordura subcutânea e visceral.

Assim como algumas questões do ANPREC, a segmentação de imagem se trata do estado do paciente e não reflete em uma ação que pode ser recomendada. Logo, ela é utilizada apenas para avaliação de risco. Importante ressaltar que os únicos escores obtidos pela segmentação são 0 e 1. Os escores 0 e 1 são obtidos, respectivamente, quando a gordura visceral ultrapassa 50% da proporção de gorduras segmentadas; e quando a gordura subcutânea é igual ou superior a 50%. Devido à falta de informações, esses valores foram escolhidos arbitrariamente.

D. Aplicação web

Ao final, foi desenvolvida uma aplicação *web* utilizando a biblioteca Python, Streamlit, especialmente voltada para a área de ciência de dados.

O paciente insere seus dados de identificação, como nome, data de nascimento, localização da neoplasia, data de diagnóstico e suas medidas antropométricas, peso atual, estatura, peso habitual na idade adulta. Em seguida, o paciente marca as opções relativas ao ANPREC e, após isso, há a opção de importar uma imagem referente à sua tomografia axial de abdômen, que será exibida sua segmentação.

Ao final, é feito o cálculo da classificação de risco e, através de uma *selectbox*, o profissional da saúde poderá escolher uma pontuação adequada para o paciente; assim, um dos algoritmos de otimização implementados será executado, ao final é exibido um plano de recomendação de ações.

E. Métricas de avaliação

1) *Segmentação de imagens*: No âmbito do processamento de imagens, o método de avaliação utilizado foi o IoU, comumente usado para avaliação de segmentação de imagens. Esta serve para quantificar o grau de sobreposição entre a área prevista por uma segmentação e a área real segmentada, fornecendo assim a medida de acurácia.

O IoU se trata da interseção da predição e da segmentação real do objeto, e a união é definida como a área total que abrange tanto a predição quanto a área real. Dessa forma, é feito o cálculo $IoU = \frac{\text{Interseção}}{\text{União}}$, resultando em um número $x \in \mathbb{R}$ tal que $x \in [0, 1]$.

2) *Planos recomendados*: Com relação aos planos gerados pelos algoritmos de otimização, coube a um profissional da saúde, da área de nutrição, avaliar o quão plausíveis são os planos gerados.

IV. RESULTADOS E DISCUSSÃO

Ao final do desenvolvimento do sistema, é preciso analisar a acurácia da metodologia de segmentação proposta, a formulação dos planos de recomendações propostos pelos métodos de otimização e a usabilidade da aplicação.

A. Segmentação de imagens

O sistema possui o objetivo de segmentar dois tipos distintos de gordura, subcutânea e visceral. Ao final do treinamento da U-Net, ao testar o modelo, foram adquiridos os seguintes resultados de acurácia, 89,84% para gordura subcutânea e 69,32% para gordura visceral. Seguem abaixo algumas imagens da segmentação:

Na Figura 2, a imagem (a) é a imagem original, enquanto a (b) é a segmentação subcutânea gerada pela U-Net.

Na Figura 3, a imagem (a) é a imagem original, enquanto a (b) é a segmentação visceral gerada pela U-Net.

No âmbito do processamento de imagens, ambas as segmentações subcutâneas e viscerais ficaram aquém do esperado. Apesar da segmentação subcutânea ter obtido uma

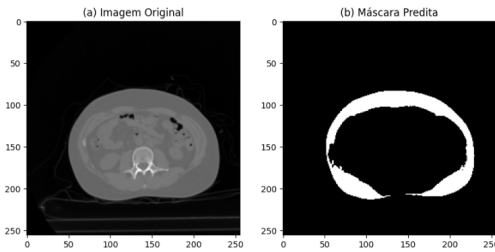


Fig. 2. Imagem de TC original e segmentação subcutânea. Elaborado pelo autor.

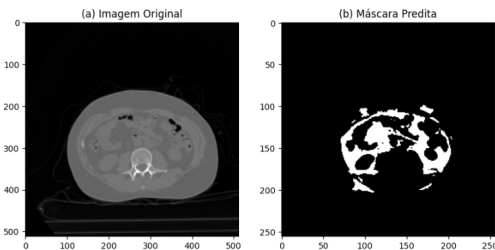


Fig. 3. Imagem de TC original e segmentação visceral. Elaborado pelo autor.

acurácia próxima a 90%, a segmentação visceral obteve um resultado inadequado.

B. Planos recomendados

Os planos gerados a partir do Algoritmo Genético, apesar de terem obtido o escore de saúde desejado e de exibir soluções rápidas, o AG apresenta falta de coesão na formulação dos planos. Isso ocorre pois o AG preza apenas por minimizar a diferença entre a pontuação de saúde atual do paciente e a pontuação desejada, sem possuir um viés de quais hábitos devem ter prioridade em ser apresentados ao usuário, levando a ajustes bruscos nas recomendações. O algoritmo apenas formula planos de recomendação que suprem o quesito de alcançar o escore desejado.

Em relação aos planos gerados pelo NSGA-II, tanto a abordagem de minimização de recomendações quanto a de minimização de saltos bruscos apresentaram ser eficazes para a formulação de planos personalizados. Isso se dá pelo fato de as implementações do NSGA-II seguirem um viés que foi explicitado pelo desenvolvedor, que dá um contexto coerente com a realidade do paciente.

A primeira versão do NSGA-II, que minimiza o número de recomendações alteradas, conseguiu reduzir o número de mudanças, mas sem necessariamente garantir uma transição gradual. Por outro lado, a segunda versão do NSGA-II, que minimiza saltos bruscos, gerou planos de recomendação mais suaves, facilitando a adaptação do paciente a uma nova rotina.

C. O sistema

Ao final, foi desenvolvido o sistema. Nele, há uma seção para o paciente adicionar suas informações pessoais e, em seguida, são apresentadas as questões de múltipla escolha do ANPREC, cuja resposta é obrigatória para seguir adiante no formulário. Antes do plano de recomendação ser gerado, o paciente tem a opção de fazer o *upload* de sua TC para o processo de segmentação. Todas as aplicações geradas possuem a mesma interface de usuário.

A interface do sistema (a), (b) e (c) apresenta a funcionalidade de solicitar aos usuários os dados pessoais, medidas antropométricas e alguns dos hábitos exigidos pelo ANPREC do paciente, enquanto na interface do sistema (d) também é apresentada a opção de fazer a importação de uma imagem de TC. Seguem abaixo imagens que ilustram a interface do sistema inteligente que utiliza o Algoritmo Genético:

GA - Avaliação de Aspectos Nutricionais na Prevenção da Recidiva do Câncer (ANPREC)

Identificação do Paciente

Nome:

Gênero:
☒ Feminino
☐ Masculino

Data de Nascimento:

Localização da doença:

Data do diagnóstico:

Fig. 4. Interface do sistema (a). Elaborado pelo autor.

Medidas Antropométricas

Peso atual (kg): - +

Peso habitual na idade adulta (kg): - +

Estatura (m): - +

Circunferência da cintura (cm): - +

Fig. 5. Interface do sistema (b). Elaborado pelo autor.

1. Tenha um peso saudável

Seu peso aumentou na idade adulta?

- ☒ Sim
☐ Não

IMC atual

- ☒ $< 18,5 \text{ kg/m}^2$ ou $> 24,9 \text{ kg/m}^2$
☐ $\geq 18,5 \text{ kg/m}^2$ e $\leq 24,9 \text{ kg/m}^2$

Circunferência da cintura

- ☒ $\geq 80 \text{ cm}$ para mulheres / $\geq 94 \text{ cm}$ para homens
☐ $< 80 \text{ cm}$ para mulheres / $< 94 \text{ cm}$ para homens

2. Seja fisicamente ativo

Você pratica atividade física diariamente?

- ☒ Não
☐ Sim

A duração e intensidade da sua atividade física diária é:

- ☒ < 45 minutos/dia de atividade moderada
☐ ≥ 45 e ≤ 60 minutos/dia de atividade moderada
☐ 60 minutos/dia de atividade moderada

Você gasta mais de 3 horas por dia com hábitos sedentários?

Fig. 6. Interface do sistema (c). Elaborado pelo autor.

8. Consumo de suplementos alimentares

Você consome suplemento alimentar? Ex.: hiperprotéico, polivitamínico, ômega 3, cálcio, vitamina D, etc.

- ☐ Sim
☒ Não

Upload de imagem para segmentação

Escolha uma imagem (png, jpg ou jpeg)

Drag and drop file here
Limit: 20MB per file • PNG, JPG, JPEG

Browse files

Fig. 7. Interface do sistema (d). Elaborado pelo autor.

Gerar Plano de Recomendação

Plano de recomendações baseado no algoritmo genético:

A duração e intensidade da sua atividade física diária é: > 60 minutos/dia de atividade moderada (Pontuação: 1)

Você consome chocolates, sorvetes industrializados, bolos com recheio e cobertura, biscoitos doces, sobremesas e doce em geral com que frequência?: Raramente ou nunca (Pontuação: 1)

Você consome algum desses alimentos com que frequência: salgadinhos fritos ou de forno, salgadinhos em pacotes, sanduíches de lanchonete, batata frita, pizzas, esfihas?: Semanalmente (Pontuação: 0.5)

Você consome cereal integral e/ou leguminosa em todas as refeições do dia com que frequência?: Diariamente (Pontuação: 1)

Você consome carne processada com que frequência?: Nunca ou raramente (Pontuação: 1)

Quantas porções de bebida alcoólica você consome diariamente?: Nenhuma (Pontuação: 1)

Você consome produtos industrializados com que frequência?: Nunca ou raramente (Pontuação: 1)

Você costuma adicionar mais sal às refeições depois de prontas?: Semanalmente (Pontuação: 0.5)

Fig. 8. Interface do sistema apresentando ações recomendadas. Elaborado pelo autor.

Ao final, são apresentadas na tela as ações que foram recomendadas pelos algoritmos.

Os três sistemas são compostos por módulos fundamentais que englobam: captura de informações do paciente, incluindo dados antropométricos e hábitos alimentares; questionário nutricional, cujas respostas são pontuadas para determinar o risco

de recidiva da doença; segmentação de imagem utilizando modelos pré-treinados para a análise de características visuais relacionadas ao estado nutricional; e algoritmo de recomendação através de estratégia de busca local, que sugere ajustes nos hábitos alimentares para atingir um escore nutricional desejado.

D. Discussão

Os resultados indicam que a modelagem de um sistema de recomendação nutricional baseada em algoritmos de otimização é um método promissor para personalizar orientações de alimentação e atividades físicas. A integração de técnicas de segmentação de imagem no sistema possibilita uma avaliação nutricional mais robusta, cruzando dados de tomografias computadorizadas com informações do questionário.

As diferentes maneiras de abordar o problema mostram que é preciso que haja equilíbrio entre a precisão do plano e o quanto o paciente vai aceitá-lo. Enquanto um Algoritmo Genético pode gerar recomendações abruptas e de difícil implementação na rotina do paciente, as abordagens multiobjetivo via NSGA-II permitem ajustes mais controlados e adaptáveis, promovendo maior aderência ao plano de recomendações.

A versão final do sistema pode ajudar tanto os profissionais de saúde quanto os pacientes, oferecendo um plano de recomendações dinâmico, ajustável e personalizado. Dessa forma, é possível aumentar a eficácia das intervenções nutricionais na prevenção da recidiva do câncer de mama.

V. CONCLUSÃO

Ao longo desta pesquisa, foram explorados conceitos fundamentais de algoritmos de busca local e do uso de Redes Neurais Convolucionais para segmentação de imagens, que quando integrados aos conhecimentos de profissionais da saúde, é possível formular um sistema robusto e pertinente na sociedade brasileira e global. Ao término do estudo, observa-se que a metodologia proposta possui pontos positivos e negativos a serem comentados.

No que diz respeito à segmentação de imagens, observou-se que a detecção de gordura subcutânea atingiu nível de precisão razoável, enquanto a segmentação visceral apresentou resultados aquém do esperado. Esse aspecto evidencia a necessidade de aprimorar o modelo de segmentação, através de refinamentos no pré-processamento das TCs ou pela adoção de uma arquitetura mais robusta.

Os experimentos demonstraram que o AG, embora capaz de minimizar a diferença entre escore atual e desejado, não foi eficiente para produzir recomendações equilibradas. Em contrapartida, as versões do NSGA-II apresentaram desempenho superior, permitindo formular recomendações de forma mais condizente com a realidade dos pacientes. A abordagem que minimiza o número de recomendações mostrou-se eficaz para

manter a aderência às recomendações sem impactar drasticamente os hábitos dos pacientes. Já a versão que minimiza saltos bruscos proporcionou transições mais suaves, evitando alterações abruptas e promovendo uma adaptação gradual. Esses resultados demonstram a necessidade do uso de técnicas multiobjetivo para equilibrar personalização e usabilidade.

A. Contribuições do Trabalho

Durante a revisão de literatura, não foram encontrados outros artigos que utilizam algoritmos de otimização para formular planos de recomendações personalizados; logo, acredita-se que este estudo possua uma abordagem ainda inexplorada na comunidade científica.

Ademais, o sistema inteligente pode contribuir não apenas para a prevenção da recidiva do câncer de mama, mas também na prevenção da primeira aparição da neoplasia, assim como efetuar planos personalizados para indivíduos que estão em fase de tratamento da doença.

B. Trabalhos Futuros

Em trabalhos futuros, há o objetivo de serem aperfeiçoadas as implementações do NSGA-II para que seja possível identificar seções em que o paciente possui baixo ou alto desempenho, para que assim seja possível formular planos ainda mais precisos. Além disso, não é descartável a possibilidade de utilizar novos métodos para geração de recomendações, como outros algoritmos de otimização ou técnicas de aprendizado de máquina, como foram abordados em alguns dos trabalhos relacionados.

Além disso, tem-se a necessidade de melhorar a seção de segmentação das tomografias computadorizadas, com o intuito de extrair informações das gorduras de forma mais eficaz.

Por fim, pretende-se disponibilizar o sistema de recomendação publicamente por meio do Streamlit Community Cloud. Dessa forma, é possível que o sistema seja acessado de qualquer lugar, proporcionando uma experiência acessível aos usuários através de acesso via URL pública.

AGRADECIMENTOS

Os autores agradecem ao Conselho Nacional de Pesquisa e Desenvolvimento Científico e Tecnológico (CNPq) pelo suporte ao desenvolvimento deste trabalho.

REFERÊNCIAS

- [1] Organização Pan-Americana da Saúde, “Câncer,” 2020, acessado em: 2 de dezembro de 2024. [Online]. Available: <https://www.paho.org/pt/topicos/cancer>
- [2] Instituto Nacional de Câncer José Alencar Gomes da Silva, “Estimativa 2018: Incidência de câncer no Brasil,” *Revista Brasileira de Cancerologia*, vol. 64, no. 1, pp. 119–120, 2018. [Online]. Available: <https://rbc.inca.gov.br/index.php/revista/article/view/115/55>
- [3] Organização Mundial da Saúde, *Marco de Aplicação da Iniciativa Mundial contra o Câncer de Mama: avaliação, fortalecimento e expansão dos serviços de detecção precoce e tratamento do câncer de mama. Resumo executivo*. Genebra: Organização Mundial da Saúde, 2023. [Online]. Available: <https://iris.who.int/bitstream/handle/10665/365790/9789240067653-spa.pdf?sequence=1>
- [4] World Health Organization, “Cancer fact sheet,” 2024. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/cancer>
- [5] —, “Obesity causes cancer and is major determinant of disability and death, warns new who report,” 2022, disponível em: <https://www.who.int/europe/news/item/03-05-2022-obesity-causes-cancer-and-is-major-determinant-of-disability-and-death--warns-new-who-report>
- [6] A. C. P. M. Pereira and M. E. L. D. Melo, “Câncer e obesidade: um alerta do inca,” *Rede Câncer*, vol. 38, pp. 34–35, 2017. [Online]. Available: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files/media/document/trc-38-artigo-cancer-e-obesidade-um-alerta-do-inca.pdf>
- [7] P. J. Goodwin, “Energy balance and cancer prognosis,” in *Cancer Prevention and Management through Exercise and Weight Control*, A. McTiernan, Ed. Boca Raton, FL: CRC Press, 2006.
- [8] Instituto Brasileiro de Geografia e Estatística, “Pesquisa do ibge mostra aumento da obesidade entre adultos,” 2020. [Online]. Available: <https://www.gov.br/pt-br/noticias/saude-e-vigilancia-sanitaria/2020/10/pesquisa-do-ibge-mostra-aumento-da-obesidade-entre-adultos>
- [9] World Cancer Research Fund and American Institute for Cancer Research, *Diet, Nutrition, Physical Activity and Cancer: a Global Perspective. Continuous Update Project Expert Report 2018*. London: World Cancer Research Fund International, 2018. [Online]. Available: <https://www.wcrf.org/wp-content/uploads/2024/11/Summary-of-Third-Expert-Report-2018.pdf>
- [10] Q. Zhang, G. Zhang, J. Lu, and D. Wu, “A framework of hybrid recommender system for personalized clinical prescription,” in *2015 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, 2015, pp. 189–195.
- [11] M. Donciu, M. Ionita, M. Dascalu, and S. Trausan-Matu, “The runner – recommender system of workout and nutrition for runners,” in *2011 13th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*, 2011, pp. 230–238.
- [12] S. Dharia, V. Jain, J. Patel, J. Vora, S. Chawla, and M. Eirinaki, “Pro-fit: A personalized fitness assistant framework,” in *Proceedings of the 28th International Conference on Software Engineering and Knowledge Engineering (SEKE)*. KSI Research, 2016, pp. 174–179.
- [13] T. Ueta, M. Iwakami, and T. Ito, “A recipe recommendation system based on automatic nutrition information extraction,” in *Knowledge Science, Engineering and Management (KSEM 2011)*, ser. Lecture Notes in Computer Science, vol. 7091. Springer, 2011, pp. 79–90.
- [14] L. Yang, C.-K. Hsieh, H. Yang, J. P. Pollak, N. Dell, S. Belongie, C. Cole, and D. Estrin, “Yum-me: A personalized nutrient-based meal recommender system,” *ACM Trans. Inf. Syst.*, vol. 36, no. 1, Jul. 2017. [Online]. Available: <https://doi.org/10.1145/3072614>