
UNIVERSIDADE FEDERAL DE OURO PRETO
DEPARTAMENTO DE CIÊNCIA DE COMPUTAÇÃO
PROCESSAMENTO DIGITAL DE IMAGENS

Projeto de Pesquisa

AGE DETECTION

Alunos:

Cesar Gabriel de Paula Batista,

Guilherme Carolino Rodrigues e Rocha,

Suleimane Ducure,

Vinícius da Silva Gomes,

Resumo

A capacidade de estimar a idade de uma pessoa através de sua foto e, com base nisso, personalizar conteúdo, oferecer produtos, restringir acesso e realizar outras personalizações baseadas na faixa etária constitui uma ferramenta poderosa para vários campos, que vão desde negócios e saúde até segurança. A detecção de idade (*age detection*) por foto, através de técnicas de processamento de imagens combinadas com métodos de Inteligência Artificial - como aprendizado de máquina e redes neurais convolucionais - permite estimar a idade ou faixa etária de uma pessoa a partir de sua face. A detecção de idade por meio de imagem facial (*age detection*) tem sido uma das pesquisas mais comuns na área da computação visual. Essa tarefa tem sido explorada extensivamente devido às suas aplicações em diversos setores, como segurança, vigilância, entretenimento, marketing e saúde.

Neste trabalho, foi desenvolvido um modelo de IA para estimar a idade de um indivíduo a partir de sua imagem facial, utilizando a arquitetura ResNet50. O modelo foi treinado com os conjuntos de dados UTK Face e FG-NET, que juntos contêm aproximadamente 18 mil imagens. A métrica utilizada para avaliar o desempenho do modelo foi o Erro Médio Absoluto (MAE). Apesar do tempo de treinamento reduzido e dos recursos computacionais limitados, os resultados obtidos demonstraram um desempenho satisfatório, considerando as limitações mencionadas.

Sumário

1	Introdução	1
2	Revisão Bibliográfica	1
2.1	Age Detection in a Surveillance Video Using Deep Learning Technique	1
2.2	Age and Gender Prediction and Validation Through Single User Images Using CNN	3
2.3	Age Estimation From Facial Image Using Convolutional Neural Network(CNN) . .	4
2.4	Human Gender and Age Detection Based on Attributes of Face	5
2.5	Age Group Classification using Convolutional Neural Network (CNN)	9
2.6	Age and Gender (Face) Recognition: A Brief Survey	11
2.7	Age and gender detection with real time database	14
2.8	Age Estimation on Low Quality Face Images	15
3	Metodologia	18
3.1	Carregamento do Modelo	18
3.2	Função de Perda e Otimização	18
3.3	Processo de Treinamento	19
3.4	Métricas Utilizadas	19
4	Resultados	19
5	Referências Bibliográficas	21

1 Introdução

Nos últimos anos, houve um crescimento significativo na pesquisa sobre extração de informações a partir de fotos de pessoas. Esses estudos têm como objetivo processar imagens para extrair informações que permitem agrupar e categorizar indivíduos com base em características como faixa etária, gênero e estado emocional. Entre essas áreas de estudo, destaca-se a detecção de idade, ou age detection, uma tarefa de visão computacional que consiste em estimar a idade de um indivíduo a partir de uma imagem digital. Com o avanço das técnicas de machine learning e deep learning, essa área ganhou notoriedade pela capacidade de automatizar o processo e alcançar maior precisão nos resultados. Segundo [Othmani et al. \(2020\)](#), as aplicações práticas da detecção de idade são amplas, que vão desde campanhas de marketing direcionadas até o monitoramento da saúde de idosos e a melhoria dos sistemas de segurança e vigilância. Nesta primeira etapa do trabalho, será realizado um levantamento bibliográfico parcial sobre os métodos e técnicas utilizados para a detecção de idade em imagens, analisar seus pontos fortes e fracos, e explorar as limitações atuais da área.

2 Revisão Bibliográfica

2.1 Age Detection in a Surveillance Video Using Deep Learning Technique

No primeiro trabalho relacionado, o artigo explora o uso de biometria suave para detecção de idade por meio de técnicas de aprendizado profundo, utilizando uma rede ResNet para a previsão de idade. Além disso, o modelo proposto é avaliado utilizando um banco de dados do Appa Real (aproximadamente 7.470 imagens) e do corpo docente de uma universidade (cerca de 40 imagens), com foco em métricas como acurácia, precisão, recall e erro médio absoluto

No segundo passo será feito o pré-processamento dos dados. Ou seja, antes mesmo do modelo ser treinado será feita uma normalização e um redimensionamento das imagens para poder garantir um padrão em todo o nosso conjunto de dados.

Em seguida, o modelo proposto utiliza uma arquitetura Wide ResNet modificada para a extração de recursos. Com essa abordagem, é possível aprender pontos-chave faciais e realizar a reconstrução de imagens utilizando técnicas como a Técnica de Reconstrução Algébrica Simultânea

(SART) para aumentar a amostragem. A arquitetura é aprimorada com uma rede de segmentação (U-Net) que emprega camadas de convolução transpostas para uma melhor representação da imagem

Finalmente, o modelo passa pela fase de treinamento. Durante essa fase, será utilizado o modelo de pontuação F1 para calcular e avaliar a eficácia do modelo, permitindo o ajuste dos parâmetros para minimizar os erros de predição.

Concluindo, será feita a avaliação do modelo para um conjunto de dados de teste para avaliar a sua precisão na avaliação da idade. O autor relata que a precisão é próxima dos 97%(Figura 1) para faixa etárias 10 a 50 e 70 a 100, e decaindo cerca de 3% para as demais, indicando a eficácia da sua metodologia. Além disso, o uso de dois bancos de dados distintos, tanto em relação à quantidade quanto aos tipos de dados utilizados, reforça a aceitação da eficácia do método proposto (Figura 2).

S.No	Dataset	Precision	Recall	<i>F</i> Measure	Accuracy
1	APPA real dataset	98.9	95.9	0.97	95
2	College dataset	93.9	96.8	0.95	93

Figura 1: Acurácia do sistema proposto. Figura adaptado do Vasavi et al. (2021)

S. No	Dataset	No of faces	No of faces correctly classified
1	APPA real dataset	7470	7153
2	College dataset	40	37

Figura 2: Resultados dos métodos propostos para os dois bancos de dados. Figura adaptado do Vasavi et al. (2021)

Contudo, o autor sugere algumas melhorias que ainda poderiam ser exploradas, como a utilização de redes residuais amplas (WRNs) para aprimorar ainda mais o desempenho. Isso envolveria o dimensionamento da largura da rede e o ajuste da profundidade para possibilitar a diminuição da reutilização de recursos.

2.2 Age and Gender Prediction and Validation Through Single User Images Using CNN

O presente trabalho discute uma nova abordagem para determinar com precisão a idade e o sexo de um indivíduo com base em uma fotografia. Dentre as principais aplicações, destacam-se a personalização de conteúdos, segurança e análise de audiência, entre outras. Os autores propõem um método que utiliza redes neurais convolucionais (CNNs) para detectar e validar a idade e o sexo por meio de uma única imagem. Para isso, eles incluíram uma camada de verificação dupla que vincula a foto ao sexo e à idade do indivíduo.

Na metodologia do trabalho, os autores usaram redes convolucionais (CNNs), que são uma classe de redes neurais projetadas para processar dados como imagens, para extrair informações da imagem facial do usuário. Em seguida, com base nas informações extraídas, foi realizada a predição da faixa etária e a determinação do gênero. A validação ocorre por meio da comparação entre os dados preditos pelo modelo e os dados fornecidos pelos usuários.

De acordo com os autores, embora os resultados obtidos na predição de sexo tenham alcançado precisão ótima, houve desafios na predição precisa da idade dos indivíduos, o que levantou a necessidade de mais refinamento e desenvolvimento para melhorar a capacidade do modelo. Outro desafio relatado pelos autores está relacionado à diversidade nas imagens dos usuários, incluindo diferenças de iluminação, ângulos e expressões faciais, que impactam significativamente o desempenho do modelo.

De acordo com os resultados obtidos neste trabalho, para a predição de gênero, o modelo apresentou uma acurácia de 82% (Figura 4), enquanto a predição de idade apresentou uma acurácia de 57% Figura (3).

	Actual		
		Male	Female
	Predicted	Male	Female
		132	101
		81	114
		Accuracy	0.57

Figura 3: Acurácia de predição de idade. Figura adaptado do [Nada et al. \(2020\)](#)

		Actual	
		Male	Female
Predicted	Male	207	26
	Female	50	145
		Accuracy	0.8224299

Figura 4: Acurácia de predição de gênero. Figura adaptado do [Nada et al. \(2020\)](#)

2.3 Age Estimation From Facial Image Using Convolutional Neural Network(CNN)

Este artigo destaca que estimar a idade humana de uma imagem é um dos problemas com mais demanda na história do campo de análise facial, além disso destaca que sistemas que estimam automaticamente idade baseado em imagens tem um grande potencial em aplicações do mundo real, se computadores forem aptos a determinar a idade do usuário então o ambiente do computador e o tipo de interação com o usuário poderiam ser ajustado a idade do usuário.

O autor propõe o processo de treinamento de um modelo de rede neural convolucional, que pode efetivamente prever a idade de faces identificadas.

Na metodologia do trabalho o autor detectou as faces das imagens presentes no seu dataset e as pre-processou, aumentou o dataset para incrementar o tamanho dos dados de treino e depois treinou o modelo da rede neural para estimar a idade das imagens faciais.

A etapa de pre-processamento consiste na detecção da face, redimensionamento da imagem e one hot encoding. Para a detecção de imagens foi usada a dlib. As imagens foram redimensionadas para 224x224 para se encaixarem com o pré requisito de input da rede neural, e também como requisito da rede neural os rótulos de output foram convertidos em vetores one hot. Para aumentar os dados as imagens foram rotadas, aumentadas, distorcidas e também algumas partes apagadas.

A etapa de treinamento consiste em um modelo de rede ResNet50 e uma camada densamente conectada por cima da Resnet50, foi retirada a ultima camada do modelo Resnet50 e adicionada uma nova camada densa de tamanho (Nx,101), e os pesos do modelo foram iniciados usando o dataset ImageNet. Após o redimensionamento, as imagens passam pela camada de entrada das Resnet50, são filtradas nas camadas convolucionais e a função de ativação ReLu é usada em cada camada convolucional. As camadas de Pooling retêm a média ou valor máximo do kernel correspondente. O modelo termina com uma camada média pooling de tamanho (Nx,2028) e a

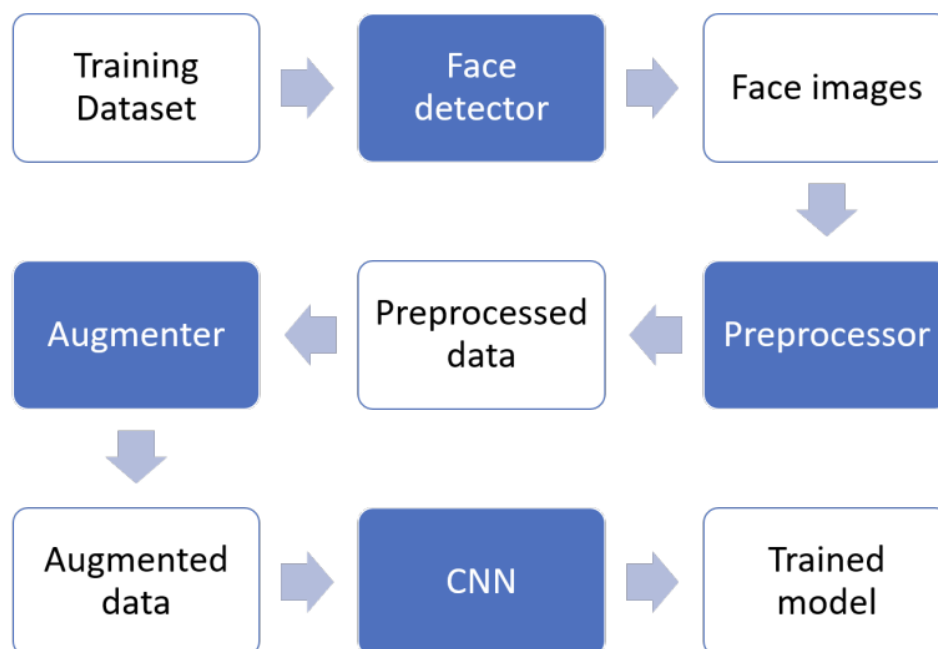


Figura 5: Diagrama do estimador de idade. Figura Retirada do [Mahjabin et al. \(2019\)](#)

saida da camada média é passada para a camada densa que foi adicionada. A entropia cruzada foi utilizada para calcular a loss.

No trabalho dois otimizadores foram testados o Adam e o SGD, após completar 20 epochs foi calculado o MAE(mean avare error) e o Cumulative Score do modelo usando o dataset de treino, usando as idades previstas e as idades reais dadas pelo dataset.

Dois datasets foram utilizados, o APPA-REAL e UTKFace na fase de treinamento, o APP-REAL contém 7,591 imagens tendo idade real e aparente associadas a ele, o UTKFace tem um dataset de faces de grande escala(20.000 imagens) com rótulos de idade, gênero, e etnia. Para teste foi usado o FG-NET que contem 1002 imagens de faces

Detalhes comparativos da MAE e da Cumulative Score de diferentes trabalhos relacionados e deste são apresentados na figura 6.

2.4 Human Gender and Age Detection Based on Attributes of Face

O seguinte trabalho descreve uma forma para detectar o sexo e a idade humanos se baseando em atributos faciais. O estudo visa desenvolver um método automático para estimar sexo e idade

TABLE II
COMPARISON OF THE METHODS FOR AUTOMATIC AGE ESTIMATION

Publications	Training datasets	Testing datasets	MAE	CS
Lanitis et al. [7]	Private (40 subject, 500 images)	A subset with 32 images	4.3	N\A
Geng et al. [8]	FG-NET (82 subject, 1002 images)	FG-NET MORPH (Subjects: N\A, 433 images)	FG-NET: 6.8 MORPH: 8.8	FG-NET: 65%, MORPH: 46%
Guo et al. [10]	UIUC-IFP-Y (1600 subjects, 8000 images)	FG-NET	5.07	47%
Chang et al. [11]	MORPH II (55,608 images with almost 3 images per person), FG-NET	FG-NET, MORPH II (Subjects: N\A, 5492)	FG-NET: 4.48 MORPH II: 6.1	FG-NET: 74.7% MORPH II: 56.5%
Guo et al. [12]	YGA (1600 subjects, 8000 images)	FG-NET	4.77	N\A
Hu Han et al. [15]	FG-NET (82, 1002) MORPH II (20569 subjects, 78207 images) PCOS4 (1802 subjects, 10036 images)	Studied on 1002, 2000 and 2200 images, respectively, from FG-NET, MORPH II, PCSO	FG-NET: 4.6 MORPH II: 4.2 PCSO: 5.1	FG-NET: 74.8% MORPH II: 72.4% PCSO: 64.0%
Proposed method	APPA-REAL (4113 images) UTKFace (23,254 images)	FG-NET (1002 images)	4.49	67.37%

Figura 6: Comparação dos métodos para estimativa de idade automática. Figura Retirada do Mah-jabin et al. (2019)

usando imagens faciais coloridas e o algoritmo Iterative Dichotomiser 3 (ID3). Abaixo é mostrado as etapas do metodo proposto (Figura 7).

Nessa metodologia, as imagens utilizadas foram extraídas do FG-NET, um conjunto que contém uma variedade de imagens faciais categorizadas em diferentes classes de idade e sexo, o que foi crucial para treinar e validar o método proposto. Após coletadas as imagens foi utilizado um algoritmo k-means para classificar as imagens em dois grupos, um contendo as imagens de homens, e o outro as imagens de mulheres.

O pre-processamento das imagens é feito antes das análises e extração das características, envolve varias etapas incluindo captura de imagens, remoção de ruído, detecção facial e normalização, para preparar imagens para análise, após isso o conjunto de dados é dividido em 80% usados para treinamento e 20% para testes, garantindo uma avaliação robusta do modelo.

Após o pré-processamento vem a parte de mineração de dados que envolve extração das características faciais das imagens, e a redução de dimensionalidade que tem como objetivo identificar e isolar as informações mais relevantes das imagens faciais, utilizando a Análise Discriminante Linear (LDA) através da técnica de Face de Fisher. Esta técnica é fundamental para reduzir a complexidade dos dados e evitar o ajuste excessivo, mantendo apenas as características mais significativas para a classificação.

A redução de dimensionalidade, por sua vez, é importante para simplificar o conjunto de dados e facilitar sua análise. Separando assim apenas os atributos que serão mais relevantes para a tarefa

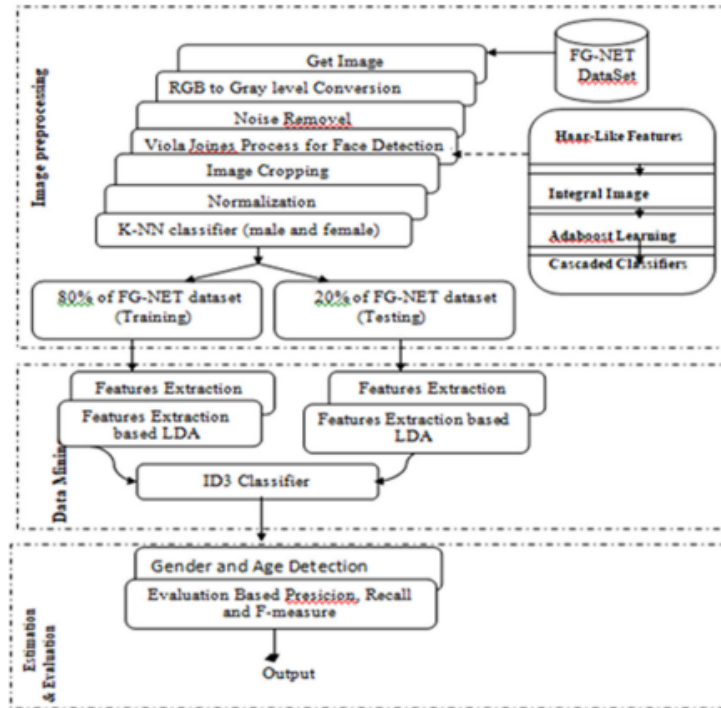


Figura 7: Etapas do método proposto. Figura adaptado do [Hameed & Farah \(2022\)](#)

de classificação ou análise. Esse processo ajuda a destacar as características mais significativas e a eliminar o ruído e a redundância, melhorando a eficiência do modelo e a precisão dos resultados. Ao reduzir o número de variáveis a serem analisadas, a redução de dimensionalidade torna a visualização e a interpretação dos dados mais manejáveis, além de otimizar o desempenho dos algoritmos de aprendizado de máquina.

De acordo com os resultados obtidos neste trabalho, o modelo apresentou uma taxa total correta de 90,93% em todas as 1090 imagens de amostra. Essa alta precisão indica que o método é eficaz na classificação de sexo e idade com base nos atributos faciais. Em sua predição de gênero, a precisão média foi de 83,066% para amostras masculinas e uma precisão média de 87,8% para amostras femininas, o que sugere que o modelo possui um bom desempenho na distinção entre rostos masculinos e femininos. Enquanto a predição de idade apresentou uma taxa total de corretas de 93,3% para as classes 1 e 2, 93,5% para as classes 3 e 4 e 86% para as classes 5 e 6, o que demonstra que esse modelo é particularmente eficaz na identificação de grupos etários mais jovens, mantendo uma precisão aceitável para os grupos mais velhos (Figura 8).

De acordo com os autores, o método proposto foi comparado com outros dois métodos, o PNN

Age	Gender	Sample Size	Correctly Detect	Correct Rate	Total Correct Rate
(3–7) (26–30)	Male	200	186	93.00%	93.30%
	Female	250	234	93.60%	
(8–13) (14–19)(20–25)	Male	210	189	90.00%	93.50%
	Female	330	322	97.00%	
(31–40)(41–50)	Male	59	52	82.00%	86.00%
	Female	41	37	90.00%	
Total	M+F	1090	1020	90.93%	90.93%

Figura 8: A taxa total de acerto na determinação da idade humana usando LDA e ID3. Figura adaptado do [Hameed & Farah \(2022\)](#)

e o SVM (Figura 9), ambos conhecidos por apresentarem altas taxas de precisão em problemas de classificação. No entanto, apesar do bom desempenho desses outros métodos, a simplicidade e a velocidade de processamento do método proposto o tornam uma escolha mais favorável para aplicações em tempo real, onde a eficiência computacional é essencial para a execução rápida e eficaz das tarefas.

Gender Type	PNN		SVM1		Proposed Method	
	Male	Female	Male	Female	Male	Female
Male	89.75	10.25	95.08	4.29	89.73	10.7
Female	11.88	88.12	4.41	95.59	12.77	87.23
Correctly classified %	88.935		95.335		88.48	

Figura 9: Resultado do teste de gênero em 20% dos itens usados para testar o conjunto de dados. Figura adaptado do [Hameed & Farah \(2022\)](#)

Além dessas métricas apresentadas foram calculados o erro médio absoluto (M.A.E) e o erro quadrático médio (R.M.S.E), que ao serem analisados sugerem que o método proposto é robusto contra valores extremos. Essa é uma vantagem sobre alguns métodos tradicionais que podem ser mais sensíveis a valores discrepantes, proporcionando assim um desempenho mais confiável em cenários práticos. Os valores são exibidos na (Figura 10).

Concluindo, podemos destacar pontos fortes e fracos dessa metodologia, sendo alguns pontos fracos que método depende de um conjunto de dados específico que pode não refletir a diversidade global, o que pode afetar a generalização dos resultados. Além disso, a seleção restrita de atributos faciais pode ignorar características importantes que poderiam melhorar a precisão. A abordagem

Accuracy measures	Result
$MAE = \frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $ where $n = 1442$	0.0198
$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$ where $n = 1442$	0.7525

Figura 10: Acurácia do método proposto. Figura adaptado do [Hameed & Farah \(2022\)](#)

simplificada de classificação etária em apenas três categorias pode levar a imprecisões, e o método pode ser sensível a variações na qualidade da imagem. Embora o algoritmo ID3 seja eficiente, a complexidade computacional, especialmente com a LDA, pode limitar sua aplicação em tempo real. A comparação com métodos existentes como PNN e SVM não aborda detalhadamente essas limitações. E como pontos fortes é que o método destaca-se por sua abordagem inovadora, combinando atributos faciais com o algoritmo Iterative Dichotomiser 3 (ID3), o que resulta em uma taxa de precisão impressionante de 90,93%. A utilização da Análise Discriminante Linear (LDA) para extração de características reforça a eficácia do modelo, ao identificar os atributos mais relevantes. A comparação com métodos existentes, como PNN e SVM, mostrou o potencial de aplicação prática em áreas como segurança e marketing que precisam de análises em tempo real, e reforçam a importância e a aplicabilidade das descobertas do artigo.

2.5 Age Group Classification using Convolutional Neural Network (CNN)

Neste quinto projeto, o autor irá utilizar um método de aprendizado ponta a ponta, ou seja, ele pretende tratar todo o sistema como uma caixa preta onde o modelo é treinado para aprender diretamente com os dados de entrada sem precisar de extração manual de recursos ou etapas de pré-processamento além do alinhamento básico e do corte de faces.

Ademais, também serão utilizadas as ferramentas FaceNet e OpenCV. O FaceNet será empregado para lidar com o desequilíbrio de classes na classificação da faixa etária. O desequilíbrio de classes ocorre quando um conjunto de dados tem uma classe com significativamente mais exemplos do que outra, resultando em uma representação distorcida dos dados, o que pode impactar negativamente a acurácia do modelo. O OpenCV será utilizado para a detecção facial em imagens, permitindo identificar a presença e a posição do rosto na imagem, o que será de grande ajuda na

classificação da faixa etária.

Na primeira etapa da metodologia, uma imagem de entrada será adicionada no pipeline para que ocorra a detecção do rosto na imagem, como mostra na Figura(11). Feito isso, opcionalmente, é possível fazer o alinhamento da face para ser ter uma melhor precisão. Após o alinhamento, a face é recortada e adicionada na rede neural para que possa fazer o seu processamento.

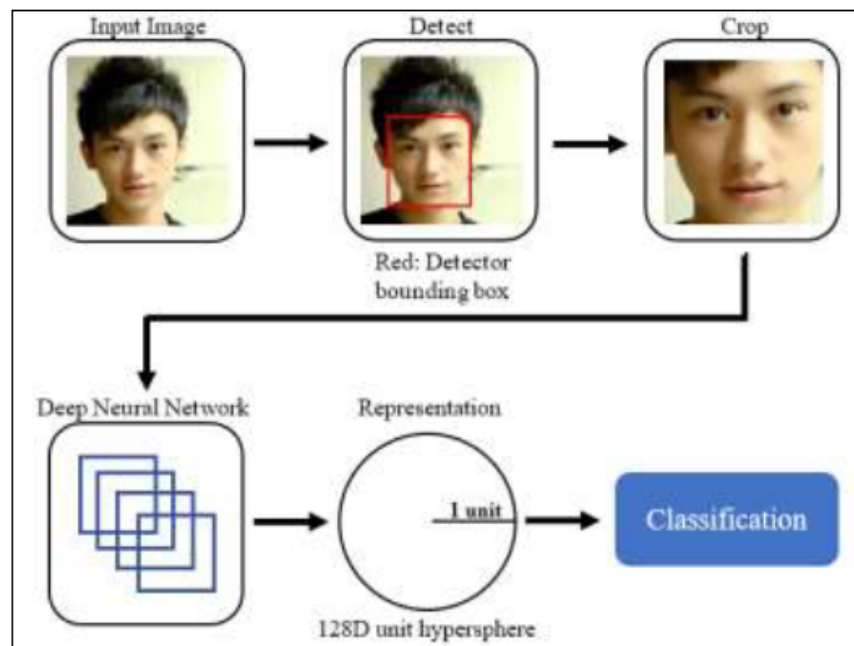


Figura 11: Processo de detecção facial. Figura adaptado do [Mustapha et al. \(2021\)](#)

Em seguida, temos o que o autor chama de 'Função de perda trigêmea', que é uma forma de otimizar o modelo por meio da incorporação de faces semelhantes (imagens âncoras e positivas) para que estejam mais próximas umas das outras, enquanto as faces diferentes (imagens negativas) estejam afastadas. Como mostra a Figura (12).

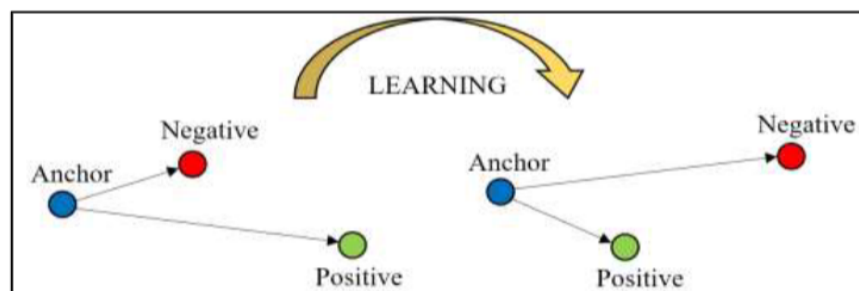


Figura 12: Função de perda trigêmea. Figura adaptado do [Mustapha et al. \(2021\)](#)

Para os experimentos, o artigo divide ele em dois tipos: O Experimento 1 foi conduzido para validar a acurácia de treinamento, enquanto o Experimento 2 foi realizado para validar a acurácia de teste. Além disso, o número de dados usados nos testes diferem entre os experimentos. Os resultados obtidos foram: para o Experimento 1 tivemos uma acurácia de 84,90%(13), e para o Experimento 2 tivemos uma acurácia de 85,12%(14).

Age group	Sample size	Correctly labelled	Correct rate	Total correct rate / Accuracy
Adolescence	849	636	74.91%	84.90%
Mature Adulthood	1614	1455	90.15%	(2091/2463)

Figura 13: Resultados do Experimento 1. Figura adaptado do [Mustapha et al. \(2021\)](#)

Age group	Sample size	Correctly labelled	Correct rate	Total correct rate / Accuracy
Adolescence	363	215	59.22%	85.12%
Mature Adulthood	692	683	98.70%	(898/1055)

Figura 14: Resultados do Experimento 2. Figura adaptado do [Mustapha et al. \(2021\)](#)

Concluindo, o modelo proposto demonstrou alta acurácia na estimativa de idade, além de lidar eficientemente com o desequilíbrio de classes, que poderia afetar a precisão do modelo. No entanto, o autor também cita algumas limitações, como a dificuldade em lidar com imagens irrestritas, o que pode resultar em perda de eficiência, além do problema de sobreajuste, especialmente quando os dados são limitados, levando à falta de generalização dos resultados.

2.6 Age and Gender (Face) Recognition: A Brief Survey

Diferente de outros trabalhos, neste estudo [Mascio et al. \(2021\)](#), os autores apresentam uma visão geral comparativa entre as abordagens mais recentes na estimativa de sexo e idade, com ênfase na introdução de novas arquiteturas de redes neurais e aprimoramento de modelos já existentes com novos componentes. A motivação por trás deste trabalho, segundo os autores, é o crescente interesse e o potencial de novos avanços na tecnologia de reconhecimento de idade e gênero, levando ao desenvolvimento de sistemas mais precisos e eficientes para o uso no mundo real. O

artigo discute tanto a introdução de novas arquiteturas de redes quanto o aprimoramento de modelos já conhecidos. Por exemplo, [Hosseini et al. \(2018\)](#) propuseram uma rede neural com 384 filtros de 3x3 pixels e duas camadas totalmente conectadas, cada uma com 256 neurônios, obtendo uma precisão de 50,7% para estimativa de idade e 86,8% para estimativa de gênero. Uma abordagem aprimorada introduziu 12 filtros de Gabor na arquitetura CNN, o que elevou a precisão para 61,3% na classificação etária e 88,9% na classificação de gênero.

AlexNet – A arquitetura básica do AlexNet (Figura 15) foi ajustada em diversas pesquisas. Alterações nas camadas de normalização e pooling, bem como a aplicação de técnicas de aumento de dados, foram implementadas para melhorar a precisão no reconhecimento de idade e gênero.

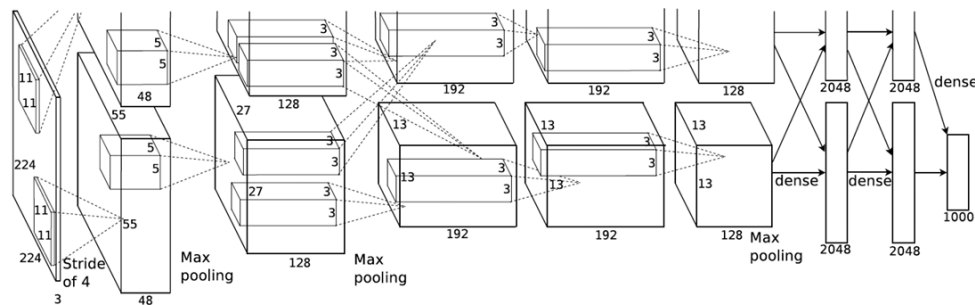


Figura 15: Arquitetura do AlexNet. [Mascio et al. \(2021\)](#)

VGGNet – Essa arquitetura (Figura 16) é frequentemente utilizada para transfer learning (aprendizado por transferência) e é considerada uma das mais promissoras devido à sua capacidade de capturar características profundas. Variações, como o VGG-Face, foram empregadas para classificar idade e gênero com maior precisão.

Outra abordagem notável foi o desenvolvimento da CNN FaceMaskNet-9, projetada especificamente para melhorar a classificação de idade em imagens faciais com máscaras, durante a pandemia de COVID-19 [Golwalkar & Mehendale \(2020\)](#). Essa arquitetura foi otimizada para detectar características faciais mesmo quando o rosto está parcialmente coberto. O FaceMaskNet-9 utiliza duas camadas convolucionais, seguidas por camadas ReLU e de max pooling, com camadas totalmente conectadas para a saída final. Técnicas de pré-processamento, como o uso do módulo DNN do OpenCV, foram integradas para melhorar o desempenho em imagens mascaradas. A base de dados utilizada foi gerada de forma amadora com técnicas de redes adversariais generativas (GAN), aumentando a diversidade das imagens de treinamento. Como resultado, o modelo alcançou uma

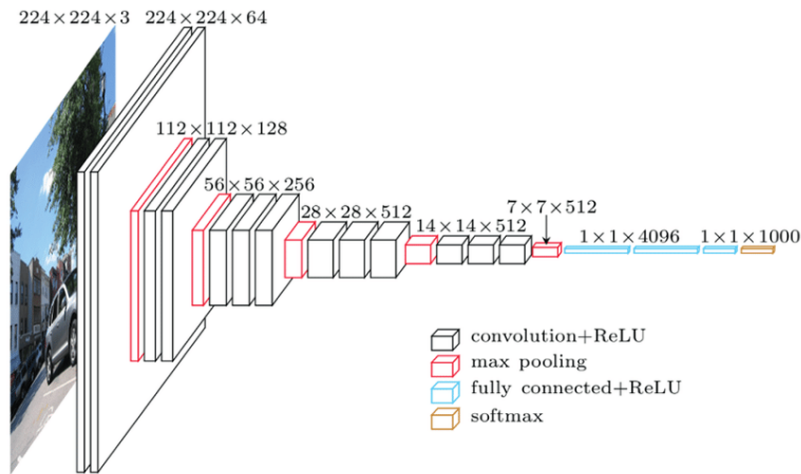


Figura 16: Arquitetura do Visual Geometry Group Network (VGGNet). Mascio et al. (2021)

precisão impressionante de 95,0

Os autores também destacam diversas dificuldades e desafios associados ao reconhecimento de idade e gênero em imagens faciais, incluindo:

- Ambientes não controlados:** Um dos maiores desafios é lidar com imagens capturadas em ambientes sem controle, onde variações de iluminação, poses, expressões faciais e oclusões afetam diretamente a precisão dos modelos.
- Bases de dados desequilibradas:** Muitas bases de dados usadas no treinamento de redes neurais são desequilibradas, contendo uma quantidade significativamente maior de imagens de pessoas jovens em comparação com idosos, o que pode gerar viés nos modelos e prejudicar a classificação correta de faixas etárias menos representadas.
- Influência de fatores étnicos e genéticos:** A classificação de gênero e idade pode ser influenciada por fatores étnicos e genéticos, que variam amplamente entre diferentes populações, criando dificuldades adicionais na construção de modelos capazes de generalizar bem para diferentes grupos demográficos.

Os autores concluem que, apesar dos avanços significativos proporcionados pelas redes neurais convolucionais profundas (CNNs) e pelo uso de grandes bases de dados, o reconhecimento automático de idade e gênero continua sendo uma tarefa desafiadora, especialmente em condições do mundo real. A pesquisa ressalta a importância de desenvolver novas arquiteturas e estratégias de aprendizado, como o aprendizado por transferência e melhorias nas técnicas de pré-processamento, para mitigar esses desafios.

Authors	Dataset	Approach	Accuracy
<i>Othmani et al.</i>	MORPH FACES FG-NET	Deep CNN (Xception, InceptionV3, ResNet-50, VGG-16 and VGG-19) 1-layer NN for regression	Age MAE: 2.52 - 4.63 yr
<i>Ozbulak et al.</i>	Adience	CNN (GilNet, AlexNet like, VGG-Face) SVM classifier Fine-Tuning	Age: 73% - 80% Gender: 87% - 95%
<i>Mallouh et al.</i>	Adience	2 CNN architectures 2 training processes for face recognition and age estimation	Age: 70% - 90%
<i>Azarmehr et al.</i>	FERET MORPH Gallagher Adience BioID PAL	MSLBP PCA LDA SVM classifier	Age: 93% - 76% Gender: 96% - 76%
<i>Duan et al.</i>	Adience MORPH	LBP+FPLBP	Age: 52% Gender: 88%

Figura 17: Resultado de aprimoração de algumas abordagens com as respectivas acurácias. Adaptado do Mascio et al. (2021)

2.7 Age and gender detection with real time database

O principal diferencial deste trabalho em relação a outros reside no uso de um banco de dados em tempo real. A pesquisa integra técnicas de visão computacional com um sistema de banco de dados em tempo real para fornecer informações precisas e imediatas sobre a demografia dos indivíduos, especificamente idade e gênero, oferecendo uma ferramenta poderosa para a tomada de decisões em diversas áreas.

Assim como em outros estudos, foi utilizado um modelo profundo de Rede Neural Convolucional (CNN), treinado com um conjunto de 18.728 imagens provenientes de vários bancos de dados, como Wiki, UTKFace e Adience. O diferencial deste trabalho está na integração com o Google Firebase, que permite o processamento, armazenamento e análise dos dados de forma instantânea. Essa capacidade de processamento em tempo real torna o sistema altamente eficiente para aplicações que requerem respostas rápidas, como vigilância, marketing personalizado e análise

demográfica.

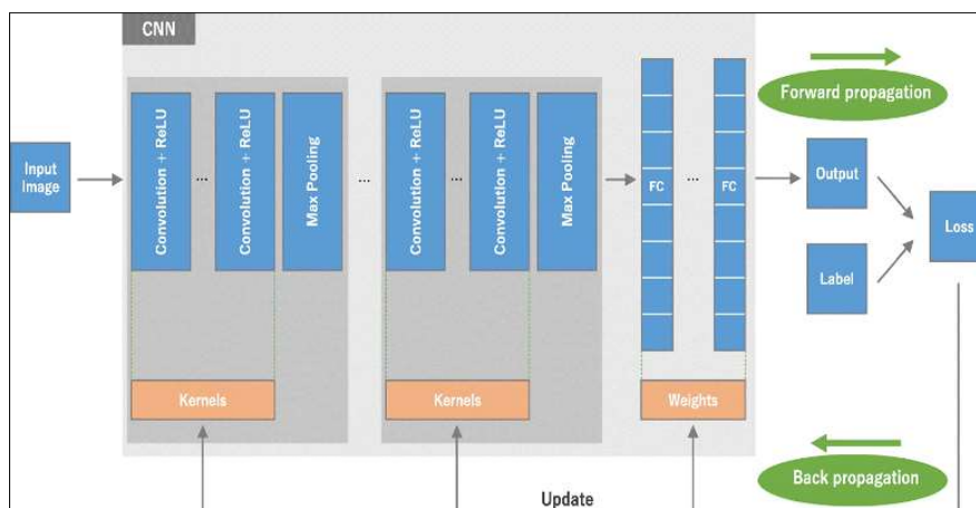


Figura 18: Arquitetura da CNN usada . Adaptado do [Bhardwaj & Sengar \(2024\)](#)

O modelo faceBox foi utilizado para identificar e capturar rostos no fluxo de vídeo. A função recebe como entrada um modelo de detecção facial (como o faceNet), o quadro de vídeo atual e uma referência ao banco de dados Firebase. O vídeo é pré-processado para que as faces sejam detectadas e demarcadas por caixas delimitadoras. Essas regiões faciais são então usadas para prever a idade e o gênero com base em modelos CNN pré-treinados, como o ageNet e o genderNet. A captura de vídeo é feita através de uma webcam, que continuamente coleta quadros a serem processados em tempo real.

Para melhorar a precisão na classificação demográfica, a metodologia incorpora técnicas que reduzem as taxas de erro, integrando múltiplas estimativas de diferentes detecções faciais. Isso ajuda a mitigar o impacto do ruído e da variabilidade no ambiente de detecção.

O artigo destaca que o método proposto demonstra desempenho superior em comparação com técnicas existentes para a classificação de idade e gênero, evidenciando a eficácia do modelo CNN e da integração com sistemas de gerenciamento de dados em tempo real.

2.8 Age Estimation on Low Quality Face Images

De acordo com os autores, ao realizarem uma revisão de literatura, perceberam que há poucos trabalhos que exploram o tema da estimação de idade a partir de imagens faciais de baixa qualidade,

tornando assim o trabalho deles relevante. Esse é um dos problemas mais frequentemente relatados na literatura, e, além disso, é comum lidar com imagens de baixa qualidade ou com ruídos em diversas aplicações. Para enfrentar esses desafios, os autores propuseram uma metodologia baseada em uma arquitetura de redes neurais convolucionais profundas recém-projetada. O estudo conta com dois conjuntos de dados de referência, o dados de IMDB-WIKI e MORPH-II.

Para a realização de trabalho, os autores apresentaram uma nova estrutura que é composta por cinco etapas:

- **Super resolução:** A primeira etapa, chamada de super-resolução, tem como objetivo melhorar a qualidade das imagens de entrada. Isso é crucial para aumentar a resolução de imagens de baixa qualidade, que frequentemente contêm ruídos ou artefatos devido a problemas de captura ou transmissão.
- **Aumento de dados:** Para facilitar o processo de treinamento das redes neurais, uma etapa de aumento de dados é incluída. Esta técnica aumenta artificialmente o tamanho do conjunto de dados de treinamento ao aplicar diversas transformações às imagens existentes, melhorando a capacidade do modelo de generalizar e criando um conjunto de dados mais robusto, especialmente útil quando há poucas amostras disponíveis.
- **Agrupamento de gênero:** A terceira etapa utiliza uma rede neural profunda para realizar o agrupamento por gênero. Essa etapa visa melhorar a precisão da estimativa de idade ao dividir as imagens faciais em dois grupos: masculino e feminino. Com essa divisão, modelos específicos de estimativa de idade são usados para cada gênero, levando em consideração que os padrões de envelhecimento podem variar entre homens e mulheres.
- **Redes profundas modificadas:** Os autores modificam duas redes profundas (Xception e DenseNet-121), integrando convoluções separáveis em profundidade. Essa modificação tem como objetivo tornar os modelos mais eficientes e melhorar a performance na estimativa de idade, considerando as diferenças nos padrões de envelhecimento entre os gêneros.
- **Procedimento de fusão:** Por fim, um procedimento de fusão é implementado para combinar os resultados das estimativas de idade de ambos os grupos de gênero. O objetivo é integrar

as previsões geradas pelas redes modificadas (Xception-DSC e DenseNet-121-DSC), aproveitando as vantagens de cada modelo e mitigando suas fraquezas, o que resulta em uma estimativa de idade mais precisa.

TABLE IV
MAE ON MORPH-II WITH IMAGE RESOLUTION 64×64

Methods	Subset	MAE(years)	MAE(years)
Xception	S1	5.8884	6.04
	S2	6.1993	
DenseNet 121	S1	7.6561	7.80
	S2	7.9593	
DSC Fusion [ours]	S1	3.9278	4.05
	S2	4.1772	

Figura 19: Erro médio absoluto com imagens de MORPH-II na resolução 64x64. Figura retirada do [Liu et al. \(2019\)](#)

TABLE I
MAE ON IMDB-WIKI WITH IMAGE RESOLUTION 32×32

Method	MAE(yaers)
Xception	10.5864
DenseNet121	11.0229
DSC Fusion [ours]	8.6750

TABLE II
MAE ON IMDB-WIKI WITH IMAGE RESOLUTION 64×64

Method	MAE(yaers)
Xception	9.8776
DenseNet121	10.2195
DSC Fusion [ours]	7.8758

Figura 20: Erro médio absoluto com imagens de IMDB-WIKI na resolução 32x32 e 64x64 . Figura retirada do [Liu et al. \(2019\)](#)

Em resumo, o artigo conclui que a estrutura proposta aborda efetivamente os desafios da estimativa de idade a partir de imagens de baixa qualidade, demonstrando melhorias significativas na precisão e fornecendo uma base para pesquisas futuras nessa área. As descobertas ressaltam a importância das técnicas de aprimoramento de imagem para obter resultados confiáveis de estimativa de idade.

TABLE III
MAE ON MORPH-II WITH IMAGE RESOLUTION 32×32

Methods	Subset	MAE(years)	MAE(years)
Xception	S1	8.4804	8.69
	S2	8.9069	
DenseNet 121	S1	10.1982	9.87
	S2	9.5421	
DSC Fusion [ours]	S1	5.0211	4.98
	S2	4.9480	

Figura 21: Erro médio absoluto com imagens de MORPH-II na resolução 32×32 . Figura retirada do [Liu et al. \(2019\)](#)

3 Metodologia

A arquitetura ResNet50 foi a rede neural convolucional escolhida devido à sua capacidade de aprender representações profundas, usando blocos residuais que permitem treinar redes muito profundas sem os problemas de desaparecimento ou explosão de gradientes. A principal vantagem dessa arquitetura é que ela consegue capturar características mais complexas ao longo das suas camadas profundas.

3.1 Carregamento do Modelo

O modelo ResNet50 pré-treinado no ImageNet foi carregado, e sua última camada linear foi substituída por uma sequência de 3 camadas lineares e dropout de 0,5: camada 1:(2048, 512), camada 2:(512,256) e camada 3(256,num_classes). O modelo ResNet50, também foi utilizado de uma outra forma, onde ele foi carregado pré-treinado porém a última camada teve apenas a saída alterada para o número de classes.

3.2 Função de Perda e Otimização

A função de perda escolhida foi o Cross EntropyLoss e o otimizador utilizado foi o SGD e Adam devido à sua eficiência em termos de tempo de convergência e por ser amplamente utilizado em problemas de visão computacional.

3.3 Processo de Treinamento

O modelo foi treinado de três formas diferentes, a primeira foi alterando apenas a saída da última camada e utilizando o otimizador SGD por 10 épocas, na segunda utilizou-se o otimizador Adam. A terceira forma a arquitetura da rede foi alterada, a última camada substituída, alterada e o treinamento utilizou o otimizador SGD por 15 épocas.

3.4 Métricas Utilizadas

Utilizou-se a métrica Mean Absolute Error (MAE) no conjunto de teste para medir a precisão da predição da idade. O MAE é amplamente utilizado, pois dá uma visão clara de quão longe, em média, o valor predito está do valor real.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

Figura 22: calculo do MAE

4 Resultados

Ao analisarmos o processo de treinamento e teste, obtivemos as seguintes conclusões:

	Modelo	Otimizador	MAE(val)	MAE (Test) FG-NET	MAE(Test) UTK FACE
Exp 1	Resnet	SGD	5.62	8,75	-
Exp 2	Resnet	ADAM	6.85	11,2	-
Exp 3	Resnet Modificada	SGD	6,05	7,2	5,8

Figura 23: Resultados obtidos

Os dados de treinamento apresentavam uma complexidade alta, incluindo imagens com múltiplas pessoas, duplicatas e uma grande variedade de classes, o que afetou a precisão do modelo.

Essas características levaram a uma rápida adaptação ao conjunto de treinamento, resultando em overfitting após aproximadamente 10 épocas.

Entre as melhorias implementadas, destacam-se o uso de um detector de faces para o pré-processamento, a exclusão de indivíduos com mais de 90 anos e a modificação da arquitetura da ResNet50. Essas alterações visaram melhorar a generalização do modelo.

O artigo de referência, que serviu como base para este trabalho, apresentou os seguintes resultados:

Otimizador	MAE
ADAM	5,3
SGD	4,49

Figura 24: Resultados obtidos pelo artigo

Os resultados obtidos no presente trabalho, apesar de não alcançarem os mesmos níveis de precisão, são satisfatórios considerando o menor tempo de treinamento e o uso de recursos computacionais mais modestos.

5 Referências Bibliográficas

Referências

- Bhardwaj, H., & Sengar, A. (2024). Age and gender detection with real time database. *International Journal of Research in Circuits, Devices and Systems*, 5, 11–16.
- Golwalkar, R., & Mehendale, N. (2020). Age detection with face mask using deep learning and facemasknet-9. Available at SSRN 3733784.
- Hameed, S., & Farah, Q. (2022, 05). Human gender and age detection based on attributes of face. *International Journal of Interactive Mobile Technologies (iJIM)*, 16, 176-190. doi: 10.3991/ijim.v16i10.30051
- Hosseini, S., Lee, S. H., Kwon, H. J., Koo, H. I., & Cho, N. I. (2018). Age and gender classification using wide convolutional neural network and gabor filter. In *2018 international workshop on advanced image technology (iwait)* (pp. 1–3).
- Liu, K.-H., Liu, H.-H., Pei, S.-C., Liu, T.-J., & Chang, C.-T. (2019). Age estimation on low quality face images. In *2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS)* (p. 295-296). doi: 10.1109/AICAS.2019.8771612
- Mahjabin, S., Alam, M. M., & Talukder, K. H. (2019). Age estimation from facial image using convolutional neural network (cnn). In *2019 international conference on computer, communication, chemical, materials and electronic engineering (ic4me2)* (pp. 1–4).
- Mascio, T. D., Fantozzi, P., Laura, L., & Ruggetti, V. (2021). Age and gender (face) recognition: A brief survey. In (p. 105-113). Springer, Cham. doi: 10.1007/978-3-030-86618-1_11
- Mustapha, M. F., Mohamad, N. M., Osman, G., & Ab Hamid, S. H. (2021). Age group classification using convolutional neural network (cnn). In *Journal of physics: Conference series* (Vol. 2084, p. 012028).
- Nada, A. A., Alajrami, E., Al-Saqqa, A. A., & Abu-Naser, S. S. (2020). Age and gender prediction and validation through single user images using cnn. *Int. J. Acad. Eng. Res.(IJAER)*, 4, 21–24.

Othmani, A., Taleb, A. R., Abdelkawy, H., & Hadid, A. (2020). Age estimation from faces using deep learning: A comparative analysis. *Computer Vision and Image Understanding*, 196, 102961-.

Vasavi, S., Vineela, P., & Raman, S. V. (2021). Age detection in a surveillance video using deep learning technique. *SN Computer Science*, 2(4), 249.