

Análise e Previsão da Série Temporal da Pandemia COVID-19 na Índia usando programação genética

Rohit Salgotra^{a,*}, Mostafa Gandomi^b, Amir H Gandomi^c

^a Departamento de ECE, Instituto Thapar de Engenharia & Tecnologia, Patiala, Índia ^b Escola de Engenharia Civil, Universidade de Teerã, Teerã, Irã ^c Faculdade de Engenharia & Tecnologia da Informação, Universidade de Tecnologia de Sydney, NSW 2007, Austrália

artigo info

Histórico do artigo:

Recebido em 17 de maio de 2020
 Aceito 26 de maio de 2020
 Disponível online 30 de maio de 2020

Keywords:

COVID-19
 Coronavírus
 SARS-CoV-2
 Previsão da série temporal
 Programação genética
 Índia

abstract

O COVID-19, declarado como uma pandemia global pela OMS, emergiu como a doença mais agressiva, impactando mais de 90% dos países do mundo. O vírus começou a partir de um único ser humano na China, agora está aumentando globalmente a uma taxa de 3% a 5% ao dia e se tornou um processo sem fim. Alguns estudos até prevêem que o vírus ficará conosco para sempre. A Índia sendo o segundo país mais populoso do mundo, também não está salva, e o vírus está se espalhando como um transmissor de nível comunitário. Por isso, torna-se muito importante analisar o possível impacto do COVID-19 na Índia e prever como ele se comportará nos dias seguintes. No presente trabalho, foram desenvolvidos modelos de previsão baseados na programação genética (GP) para casos confirmados (CC) e óbito (DC) em três maioria estados afetados, ou seja, Maharashtra, Gujarat e Delhi, bem como toda a Índia. Os modelos de previsão propostos são apresentados por meio de fórmula explícita, e a importância das variáveis de predição são estudadas. Aqui, parâmetros estatísticos e métricas têm sido utilizados para avaliação e validação dos modelos evoluídos. A partir dos resultados, descobriu-se que os modelos baseados em GP propostos usam funções simples de linkage e são altamente confiáveis para a previsão de séries temporais de CASOS COVID-19 na Índia.

© 2020 Elsevier Ltd. Todos os direitos reservados.

1. Introdução

Síndrome Respiratória Aguda Grave Coronavirus 2 (SARS-CoV-2) ou doença coronavírus 2019 (COVID-19) é um surto global de pandemia, o mundo enfrenta hoje. O vírus, que migrou de morcegos para humanos, originários de Wuhan, capital da província de Hubei, na China, influenciou mais de 180 países do mundo. O primeiro paciente foi notificado em 8 de dezembro de 2019, em Wuhan, pela administração chinesa [1]. Um mês depois, a primeira morte pelo vírus foi registrada em 9 de janeiro de 2020. No mesmo dia, a Organização Mundial da Saúde (OMS) declarou que uma nova doença do coronavírus foi identificada e está se expandindo multilateralmente todos os dias [2]. O vírus que começou a partir da China, migrou para o terceiro mundo e o primeiro caso na Tailândia foi relatado em 13 de janeiro de 2020 [3]. As autoridades chinesas tentaram conter o vírus impondo certas ações rigorosas, incluindo fechamentos de aeroportos, fechamentos de vias altas, interrupções ferroviárias, reunião pública suspensão, paralisação do transporte público, fechamento de lojas, atividades em massa e qualquer outra atividade que contabilizasse encontros sociais foram interrompidos imediatamente. Todas essas medidas onde empregadas para

<https://doi.org/10.1016/j.chaos.2020.109945> 0960-0779/©
 2020 Elsevier Ltd. Todos os direitos reservados.

minimizar o efeito da transmissão do nível comunitário da desviá-facilidade [4]. As autoridades chinesas assumiram o controle da situação e coletaram dados da International Air Transport Association (IATA) de 2018 para identificar e verificar o índice de vulnerabilidade de doenças infecciosas (IDVIs) em novos países onde o vírus pode ter transmitido fora da China [5]. Deve-se notar que o IDVI tem uma faixa de [0, 1], maior o valor do IDVI, menor é o risco de transmissão e vulnerabilidade da doença. O vírus afetou mais de 85,000 população chinesa e os destinos iniciais foram Hong Kong, Bangkok, Tóquio e Taipei, todos tendo um IDVI acima de 0,65 [6].

Embora inúmeros esforços tenham sido realizados, mas ainda assim o vírus não foi controlado e até 19 de janeiro de 2020, inúmeros casos em todo o mundo foram relatados [3]. A OMS declarou a desamenidade como uma situação de emergência para todo o mundo em 31 de janeiro de 2020 e, em 11 de março de 2020, foi declarada como uma nova ameaça de pandemia global [4]. Em 12 de maio de 2020, quase 4.006.257 foram relatados em todo o mundo com uma contagem total de óbitos (DC) de 278.892, totalizando um aumento diário de 26% e 28% em casos confirmados (CC) para óbitos por dia [6]. O país mais afetado é os EUA, o segundo país mais afetado é a Federação Russa, seguido pelo Reino Unido, Espanha, Itália, Alemanha, França e Turquia. O vírus que começou a partir da China, engoliu quase todos os países do mundo, sendo a região mais afetada o continente europeu [7]. Estimativas médias

* Autor Correspondente.

Endereço de e-mail: r.03dec@gmail.com (R. Salgotra).

também foram traçadas para calibrar e projetar modelos de transmissão COVID-19, antes que novas investigações e controle pandemia mea- certas possam ser implementado [8]. Também pode-se notar que o vírus começou a partir de um único indivíduo, mas migrou para o nível de cluster e na situação atual, é enormemente aumenta como um sistema de transmissão de nível comunitário [11].

Índia é o segundo país mais populoso do mundo com 1,3 bilhão de pessoas para servir, tendo uma renda familiar média classificada em 112º de 164 países pelo Banco Mundial e com o 150º lugar na área de saúde global pelo Fórum Econômico Mundial. Esta condição crítica estava sob o scanner de todo o mundo, quando a pandemia COVID-19 entou pela primeira vez em solo indiano [9]. O primeiro caso foi relatado em 30 de janeiro de 2020 e esperava-se que a Índia não pudesse sobreviver ao calor e devido à falta de serviços essenciais, a vida de milhões de pessoas estará em jogo. Seus mais de três meses desde o primeiro caso e o total de casos confirmados atingiram um nível de 70756 a partir de 12º Maio de 2020. Com uma taxa de recuperação de 31,7%, um total de 22455 pessoas foram recuperados e liberados, enquanto 2293 pessoas foram resgatadas após a morte [10]. Se compararmos os mesmos números com o terceiro mundo, os EUA têm um número total de casos confirmados, totalizando 1.271.645 com taxa de aumento de 25,0 00 pessoas por dia e uma contagem de óbitos de 76.916 no mesmo período. Espanha com um total de 224.390 casos confirmados contra 26.621 mortes, Federação Russa com 221.344 casos confirmados contra 2009 casos de morte e um aumento diário de mais de 10,0 00 casos. Reino Unido está em quarto lugar com 219.187 casos confirmados contra 33.854 mortes, e a China, onde a pandemia começou, soma 84.450 casos confirmados contra uma contagem de mortes de 4643 pessoas [6]. A principal razão para tão pouco efeito do COVID-19 na Índia, deve-se à resposta oportuna dos respectivos governos centrais e estaduais. Desde o primeiro dia do surto na Índia, o governo indiano tem escaneado cada pessoa que vem da China para o país e pessoas que tiveram qualquer histórico de viagens chinesas nos últimos dias. O primeiro bloqueio nacional por 21 dias foi anunciado pelo governo em 23 de março de 2020 e foi ainda mais estendido por outros 21 dias até 3 de maio de 2020, que foi prorrogado até 17 de maio de 2020 (podendo ser prorrogado até 30 de maio de 2020). As medidas do importante incluíram resposta oportuna para fornecer unidades de saúde, rastreamento de contact, testes extensivos, mobilização comunitária e outros ajudaram para conter o vírus e manter uma baixa taxa de mortalidade. Diferentes estados já se recuperaram de vários anúncios e que os ajudaram a manter um controle sobre a nova situação. Odisha, Kerela e Tamil Nadu tem uma longa história de desastres natural e medidas de precaução já foram tomadas pelo governo. Maharashtra em um todo usa drones para monitorar distanciamento social e bloqueio. Uma estratégia de contenção de clusters também foi empregada para diagnosticar e conter o vírus. Isso tem sido feito por meio de levantamento, detecção e rastreamento de contato de cerca de 3 km de área onde mais de três pacientes são diagnosticados [9].

O efeito potencial do COVID-19 levou vários ganhos sobre as características do coronavírus e um grande número de estudos estão em processamento para estimar o possível devastação pelo vírus e para derivar uma vacina para sua cura. Também foi descoberto que o vírus tem um efeito adverso em idosos, bem como em pessoas que sofrem de algum tipo de dificuldade tais ataques cardíacos, doenças respiratórias, e outros, e é uma grande preocupação para as autoridades manter um controle sobre o vírus para que o dano mínimo possa ser feito [12]. Vários estudos têm sido conduzidos por pesquisadores de todo o mundo para estimar o possível impacto do coronavírus. Os principais estudos incluem simulações estocásticas [13], modelo de distribuição weibull [14], modelo de crescimento exponencial [16], distribuição lognormal [15] e outros [17]. Os ganhos foram capazes de prever um período médio de incubação de 5,1 dias e um total de 14 dias de quarentena necessários para a análise do vírus dentro de uma pessoa [13]. Mas nenhum desses estudos poderia estimar a taxa exata de reprodução e, portanto, não foi feito muito para pre-dict como o vírus vai efeito nas próximas semanas ou assim. Além disso, todos os estudos foram feitos sobre a China e pouco trabalho foi feito em relação ao subcontinente indiano.

No presente trabalho, um novo modelo baseado em programação genética (GP) [18] para a previsão da série de tempos dos cenários COVID-19 na Índia foi proposto para estimar a possível disseminação de o vírus. O conjunto de dados para avaliação é retirado a partir de [19]. GP é uma versão aprimorada do algoritmo genético (GA) [20], na qual novas soluções são geradas como programas baseados em computador em vez de simples strings binárias [21]. GP minério mais precisamente programação de expressão genética (GEP), é a versão mais recente do GP e tem sido analisado por vários pesquisadores para fazer modelos de predição, lineares modelos de regressão e outros [22-24]. O modelo GEP tem sido usado para pre-dict o número total de casos na Índia com base em dois grandes parâmetros, estes incluem confirmados casos (CC) e contagem de óbitos (DC). Note-se que essas séries temporizadas foram desenvolvidas a partir da data do primeiro bloqueio do AF- Ter, que é a partir de 24 de março de 2020. Além disso, o conjunto combinado de dados utilizado no presente trabalho está disponível em. Uma razão maior para usar a modelagem GEP é que essa abordagem é mais eficiente em comparação com as técnicas clássicas e são mais estáveis em comparação com redes neurais artificiais. Além disso, modelos baseados em GEP geram equações de predição simples que podem ser otimizadas de acordo com os requisitos do usuário final. Outra razão para o uso do GEP, é que esses modelos não precisam de informações prévias para desenvolver equações de previsão. No geral, pode-se dizer que os modelos de previsão propostos pela modeling baseada em GEP têm melhor calibração e podem analisar os resultados de forma muito eficaz [25,26]. Assim, no presente trabalho, os modelos GEP são propostos com base nos dados brutos extraídos de fontes autênticas desde 24 de março de 2020.

O artigo é organizado em 4 seções, a primeira seção inclui a introdução como discutido acima. No seção 2, preliminares técnicas e análise de modelos são apresentados, fornecendo detalhes sobre os básicos do GEP e modelo GEP proposto. A seção 3, fornece os resultados e a discussão relacionados a diferentes cenários do COVID-19. Aqui deve-se notar que três estados mais afetados do Índia foram levados em consideração e os modelos GEP para todos os estados foram propostos. Estes estados incluem Maharashtra, Gujarat e Delhi. Esta seção inclui ainda o modelo validação, estudo comparativo e importância variável de todos os componentes, para todos os estados, que são necessários para o desempenho preciso do modelo GEP proposto. Além disso, também foi discutida a análise em relação à unidade percentual, efeito da temperatura e variáveis no COVID-19. Finalmente, conclusões perspicazes e recomendações futuras são desenhadas na seção final.

2. Preliminares Técnicas e Calibração de Modelos

GEP é um algoritmo evolutivo altamente eficaz e provou seu valor em comparação com o GA. O algoritmo produz novas equações em vez de cordas binárias e, portanto, tem o avanço direto de formulações matemáticas para problemas dimensionais mais elevados que caso contrário, não é possível com um GA padrão. Para formular o modelo GEP para a Índia, é realmente muito importante investigar os modelos existentes e analisar se o gep modelos proposto será significativo o suficiente ou não. Vários modelos como Ace-Mod (Australian Census-based Epidemic Model) [27], neural network models [28] e outros foram empregados para acessar o situação e fornecer previsões exatas. Embora esses modelos sejam um pouco significativos, mas o primeiro modelo AceMod tem sido usado para previsão de influenza [27] e tem pouca relevância para o COVID-19. O outro modelo baseado em rede neural usa o método de memória de longo prazo raso (LSTM), juntamente com o modelo baseado em regras difusas para prever o cenário atual. Foram encontrados valores de erro quadrado médio muito alto (RMSE) e correlação (R^2) tornando o modelo iluminado, vulnerável a incertezas. Ambos os métodos discutidos são básicos e discretos por natureza e requerem mais conjuntos de dados para fornecer previsões exatas. Além disso, são técnicas clássicas e representam uma implementação muito desafiadora quando comparadas à modelagem

baseada em SIM-ple GEP. O GEP permite que um sistema seja calibrado facilmente e até mesmo prever uma solução precisa, bem como confiável sob restrições mínimas [25]. No presente trabalho, duas grandes séries temporais, incluindo CC e DC, foram levadas em consideração para acessar e prever o possível impacto do COVID-19 nos principais estados em toda a Índia, bem como na Índia em um todo. Os estados que são levados em consideração são Maharashtra com o maior número de 23.401 CC contra 868 DC, Gujarat com 8541 CC contra 513 DC e Delhi com 7233 CC contra 73 DC. Metodologia detalhada para os modelos baseados em GEP apresentados em subseções subsequentes.

2.1. Programação de Expressão Genética

GP em um todo é uma extensão de GA e é baseado no mesmo princípio da teoria de Darwin de seleções naturais ou sobrevivência do mais apto. Aqui novas equações ou simplesmente modelos de programa baseados em computador são criados para encontrar uma relação entre as variáveis de entrada e saída. É basicamente um programa baseado em computador e cria uma estrutura semelhante a uma árvore comumente referida como modelo GP baseado em btree e são declarados em um programa funcional - linguagem ming [22]. No geral, o GP é uma estrutura hierárquica com funções e terminais. A versão mais recente da modelagem GP é o GEP, que usa uma sequência de caracteres de comprimento fixo em vez de representação de árvore clássica de um GP modelo. A nova estrutura consiste em cinco componentes principais, como conjunto de funções, parâmetros de controle, condição terminal, função de aptidão e conjunto terminal. Todos esses componentes formam coletivamente uma simples árvore parada é conhecida como uma árvore de expressão (ET). A maior vantagem de usar esse tipo de metodologia é que ela é extremamente simples e funciona em nível cromato. Também consiste em propriedades multigênicas e pode ser usada para a evolução de modelos de subprogramação complexos e não lineares [29]. Cada modelo GEP consiste em uma lista de símbolos de comprimento fixo, um conjunto de funções (por exemplo, +, -, ×, /, Log) e um conjunto de terminais (por exemplo, a, b, c, 3). Assim, em termos de conjunto terminal e conjunto de funções, um GEP pode ser uma invenção de cromossomos múltiplos que são capazes de ressentir-se na forma de qualquer parse árvore. Para decodificar esta informação, a linguagem Karva é usada no nível do cromossomo [30] e um gene simples na língua Karva é dado por

$$/Log ++ c3 \text{ a partir de} \quad (1)$$

onde a , b e c são variáveis e 3 é uma constante. O expression em 1 é chamado como uma expressão K ou geralmente uma notação karva. O modelo formulado acima pode então ser evoluído na forma de um ETs simples, conforme dado pela Figura 1. A expressão na Equação 1, pode ser convertida em uma expressão k. Esta expressão é a raiz

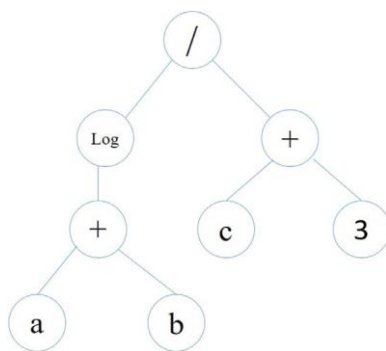


Fig. 1. Representação de um ET.

matemáticas complexas [31]. Assim, uma expressão k simplificada é apresentada na forma de equações matemáticas como dado por

$$Log(a + b) / (c + 3) \quad (2)$$

Observe que a expressão k acima discutida pode-se estimar que o comprimento total dos genes em um modelo GEP permanece o mesmo, mas o número total de ETs continua mudando com respeito ao problema em questão [22]. O modelo GEP, assim, formulado ainda mais descobre que certos elementos redundantes estão presentes na notação que não são significativos para o mapeamento genético. Assim, para que um modelo GEP seja um modelo eficiente, o comprimento total de uma expressão k deve ser igual ou inferior ao comprimento total do gene GEP. Aqui deve-se notar que uma metodologia aleatória de cauda de cabeça é seguida por um GEP para selecionar um gene. A cabeça pode ter o símbolo de função e o símbolo do terminal, mas a cauda tem apenas um símbolo terminal [22].

Para uma população uniformemente distribuída aleatoriamente inicializada, um modelo GEP consiste em cromossomo de comprimento fixo para cada membro da população. O próximo passo é avaliar a aptidão do cromossomo conforme o requisito do problema, em seguida, avaliar a solução usando a seleção da roleta com elitismo e, finalmente, encontrar a melhor solução baseada no fitness para reproduzir novos indivíduos com certas modificações. Observe que um critério de rescisão, em termos de número de geração ou algum valor de erro aceitável, também é definido. Aqui, a solução final obtida após o término dos critérios de terminação, é considerada como a possível solução do problema em teste. O diagrama esquemático para as etapas fundamentais de um modelo GEP é dado pela Figura 2. O algoritmo devido à presença do mecanismo de seleção de rublo wheel, é muito eficiente na otimização e clonagem do melhor indivíduo com respeito à mudança de iterações e, finalmente, encontrar a melhor solução [25].

2.2. Modelo GEP proposto

Para ter uma compreensão clara do número total de casos DE COVID-19 em toda a Índia, dois parâmetros principais são levados para a eração consídua, incluindo CC e DC. Ambos os parâmetros são tomados para acessar com precisão e prever o efeito do COVID-19. Para avaliação de desempenho, oito registros anteriores são usados na série temporal e o melhor modelo GEP baseado neles é selecionado. O conjunto numérico de dados utilizado é dividido em dois sub-conjuntos de dados e são equivalentemente utilizados para treinamento, bem como fase de teste/validação. Além disso, é um fato bem conhecido que o desempenho de um algoritmo de evolução não pode ser julgado usando uma única execução e, portanto, múltiplas corridas do conjunto de dados foi realizado para reduzir o erro e prever uma saída quase ótima [25]. Aqui várias corridas do mesmo conjunto de dados foram simuladas, ajudando assim o algoritmo na saída exata provendo mesmo que as instâncias totais para a experimentação é limitada em número. Como etapa de avaliação adicional, 70% dos dados foram utilizados para a realização de tarefas de treinamento e 30% de descanso foram utilizados para a realização de testes/validação. Aqui deve-se notar ainda que os dados de treinamento utilizam evolução genética e o melhor modelo é previsto usando coeficiente de correlação. Assim, um novo modelo foi proposto para um melhor desempenho para o treinamento e pode de alguma forma trabalhar bem o suficiente para a fase de testes.

Além disso, o modelo GEP é muito afetado pela escolha do parâmetro. A fim de ter um modelo justo várias corridas foram conduzidas para encontrar solução global, alterando as configurações paramétricas. A configuração paramétrica inicial foi baseada no modelo anteriormente introduzido, conforme dado por [22]. Além disso, para o melhor desempenho, a função de condicionamento físico é evoluída em relação ao erro quadrado médio (MSE) e a nova função de aptidão pode ser matemática-

de ET, que lê através do nó funcional e, finalmente, para o nó terminal. Esse tipo de interpretação permite uma derstanding mais rápida das complexidades

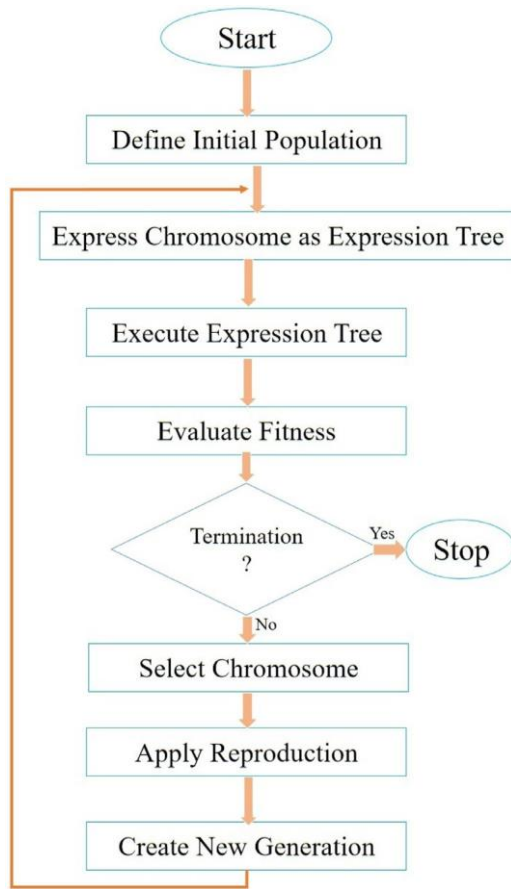


Fig. 2. Representação de um algoritmo GEP.

as métricas de erro quadrado de média raiz (*RMSE*) e coeficiente de correlação (*R*) são levadas em consideração para avaliar

Tabela 1

Configurações de parâmetro para algoritmo GEP.

Parâmetro	Configurações
Geral	
Cromossomo	30
Gene	5
Tamanho DC	5
Tamanho da cabeça	4
Tamanho da cauda	5
Tamanho do gene	14
Função de vinculação	Adição/Mínimo
Taxa de mutação do operador genético	+ , × , ÷ , √ 0.00206
Taxa de inversão	0.00546
Taxa de transposição IS e RIS	0.00546
Taxa de recombinação de um ponto e dois pontos	0.00277
Taxa de recombinação e transposição genética	0.00277
Constantes numéricas por gene	
	10
Tipo de dados	Ponto flutuante
Gama	[-10, 10]

o modelo, que são calculados como

$$= \frac{\sum_{i=1}^n |h_i - \tilde{h}_i|}{n} \quad RMSE \quad (4)$$

$$= \frac{\sum_{i=1}^n (h_i - \tilde{h}_i)(t_i - \tilde{t}_i)}{\sqrt{\sum_{i=1}^n (h_i - \tilde{h}_i)^2 \sum_{i=1}^n (t_i - \tilde{t}_i)^2}} \quad R \quad (5)$$

onde n é o número total de amostras, h_i e t_i são as saídas reais e pretendidas, \tilde{h}_i e \tilde{t}_i são médias do real e pretendido saídas para a saída i th. Além disso, é um fato bem conhecido que apenas os valores R não podem ser considerados como métricas de boa avaliação. Isso ocorre porque os valores R não mudam significativamente mudando a saída de qualquer modelo preditivo. Portanto, há a exigência de alguns outros parâmetros ou indicadores que podem avaliar o algoritmo proposto. No presente trabalho, a *RMSE* também foi levada em consideração. Aqui o *RMSE* é uma função de erro e valores mais baixos deste parâmetro indicam que um modelo mais preciso pode ser elaborado. Além disso, Smith et. al, [32] afirmou que para que um modelo seja confiável e preciso, o coeficiente de correlação entre os valores desejados e pretendidos deve ser forte. Assim, pode-se dizer que qualquer modelo com menor valor de *RMSE* e valores mais elevados de R tem a capacidade de fornecer séries temporárias confiáveis previsões [33].

Para validar externamente o modelo GEP, também foram empregados critérios utilizados por [34]. A principal característica deste critério é que as encostas de regressão (k ou k') devem estar próximas de 1 e devem estar em torno da origem. O valor dos parâmetros n e m deve ser inferior a 0,1, enquanto a previsibilidade externa R_m deve ser superior a 0,5 [35]. Além disso, o coeficiente de correlação quadrado (Ro^2) e o coeficiente (Ro^2) devem estar próximos de 1. Aqui deve-se notar que o valor de Ro^2 está entre os valores previstos e desejados, onde como Ro^2 está entre valores desejados e previstos re-especificamente [25]. Mais detalhes sobre outros parâmetros para validação externa são dados na Tabela 2. Todos esses parâmetros desempenham uma parte importante em garantir uma boa probabilidade de previsão de cada modelo proposto e também analisar a forte validade do modelo.

3.1. Modelo GEP para toda a Índia

Uma comparação dos valores reais ou previstos para CC e DC são dadas na Figura 3. As formulações matemáticas desagregadas acima representam uma complexa organização de constantes, operadores e variáveis para prever

matematicamente formulado como

$$Fitness = \frac{(1 + MSE)}{1} \times 100 \quad (3)$$

A função de aptidão formulada na equação 3, é usada para todos os casos sob investigação. Uma discussão detalhada sobre a validade do modelo e o estudo comparativo é apresentada em seção consecutiva. **3. Resultados Numéricos e Estudo Comparativo**

O algoritmo GEP para elaborar novo modelo para COVID-19 no dia foi implementado utilizando a GeneXpro Tool [31]. Para os testes de ópera genético, as configurações dos parâmetros são dadas na Tabela 1. O algoritmo foi executado para um determinado conjunto de parâmetros e as simulações foram formadas até que nenhuma melhora adicional possa ser notada no modelo GEP. Aqui o tamanho da cabeça e a contagem total de genes ajuda na evolução da arquitetura geral do modelo GEP. A contagem genética para cada cromossomo no modelo determina o número total de termos no modelo GEP e cada gene corresponde ainda mais a cada sub-ET. No presente caso, os níveis totais ideais para o tamanho da cabeça e genes são tomados como 4 e 5, respectivamente. Para tamanho genético maior que 1, modelos matemáticos podem ser ligados usando funções de linkage. Além disso, as funções de ligação utilizadas no trabalho atual são muito básicas na formulação e, portanto, um modelo GEP simples, mas ideal, foi elaborado. Os parâmetros estatísticos para o modelo GEP proposto são dados na Tabela 2. A próxima seção detalha os resultados e a discussão correspondentes ao modelo GEP proposto.

Aqui são elaborados dois modelos baseados em CC e DC e as configurações parâmetros para sua avaliação são dadas na Tabela 2. Aqui deve-se notar que

a produção. A partir da [Figura 3](#), é evidente que ambos os modelos de predição dão resultados quase equivalentes aos do CC original e dc. Além disso, a partir dos modelos, pode-se ver que até 13 de maio de 2020, há número total de CC e DC

Fig. 3. Casos experimentais versus previstos para COVID-19 na Índia usando o modelo GEP.

é em torno de 80,0 0 e 2500, respectivamente. Os valores originais ver-sus previstos para o mesmo período são quase semelhantes. A **Figura 3**, também mostra as saídas previstas para esses modelos e constatou-se que nos próximos 10 dias isso é até 23 de maio de 2020,

3.1.2. Importância Variável

As variáveis preditoras são uma parte importante e integral de um modelo GEP [36]. Esses parâmetros ajudam na compreensão da contribuição de todas as variáveis do modelo. Aqui é seguido um fenômeno de randomização para cada valor de entrada, a fim de analisar a importância de cada variável e, em seguida, encontrar a redução média em R^2 entre os valor previsto e a saída desejada. Os resultados obtidos para todas as variáveis de previsão são normalizados tais

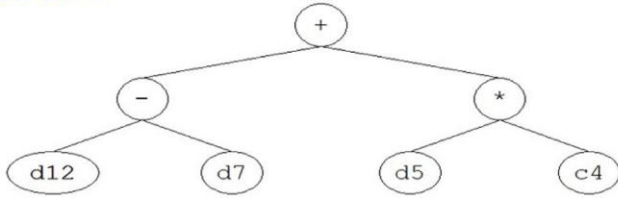
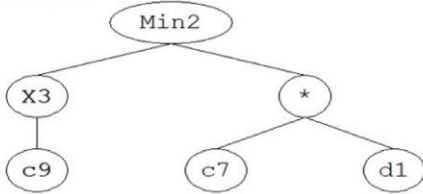
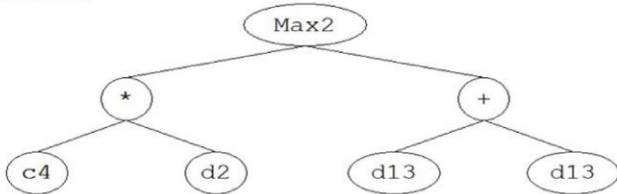
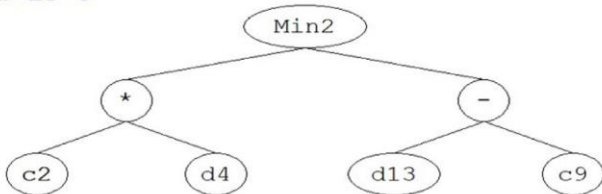
$$\begin{aligned} G2C8 &= 718,745292468032; \\ G3C5 &= -11,1531788659534; \\ G4C5 &= 18,5318425566417; \\ G4C7 &= 3,4333220229194; \quad y = 0 \\ &. 0; \end{aligned}$$

$$\begin{aligned} y &= \min ((d(14) + G1C1), (d(10) + d(10))); \quad y = \\ &(y + (d(2) / (G2C8 - d(4)))) / 2 . 0; \quad y = (y + \min \\ &(((d(2))^2), (G3C5^2))) / 2 . 0; \quad y = (y + (((G4C \\ &5 + d(1)) / 2 . 0) + (G4C7 * \end{aligned}$$

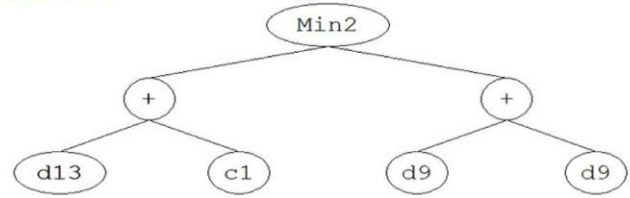
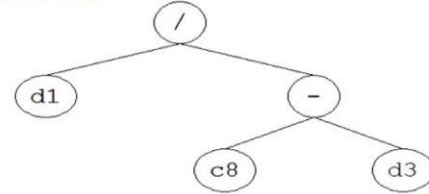
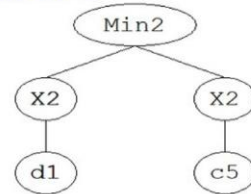
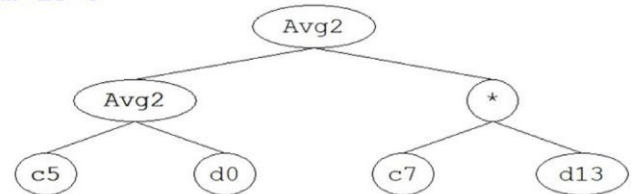
$$d(14)) / 2 . 0)) / 2 . 0;$$

Resultado=y;

Fim

Sub-ET 1**Sub-ET 2****Sub-ET 3****Sub-ET 4**

(a) ETs for Confirmed Cases

Sub-ET 1**Sub-ET 2****Sub-ET 3****Sub-ET 4**

(b) ETs for Death Cases

Fig. 4. Árvores de expressão (ETs) para a modelagem de COVID-19 na Índia.

Algoritmo 2 Modelo de previsão da Série Hora gerado para DC em toda a Índia.

função Resultado=GEPModel(d) G1C1 =
12,0 0 0142754615;

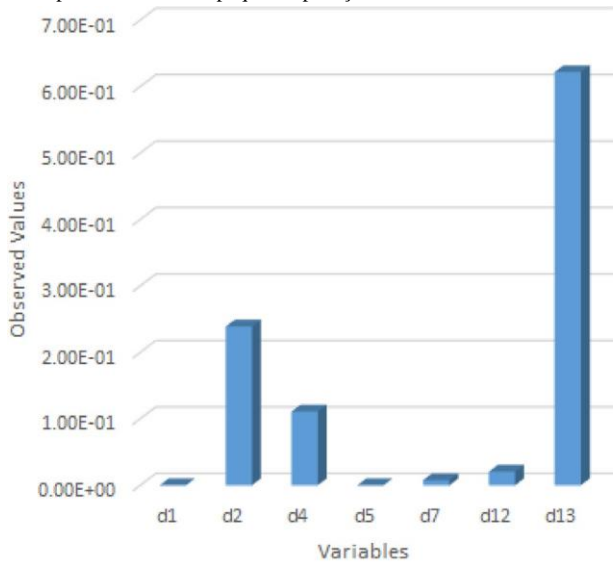
que a adição de todas as variáveis equivale a 1. A partir dos resultados na **Figura 5**, pode-se ver que para toda a Índia a partir de 13 de maio de 2020, a variável $d13$ afeta muito os algoritmos e é o a variável im-portant tanto para

CC quanto para DC. No caso de CC, o modelo é altamente sensível a duas outras variáveis: d_2 e d_4 . Nas próximas subseções, três grandes estados da Índia também foram estudados e desempenho de modelos GEP propostos de CC e DC para esses três estados tem foram analisados. Observe que os detalhes básicos sobre os ETs, a importância variável e os valores estatísticos não foram novamente re-feridos para evitar a repetição.

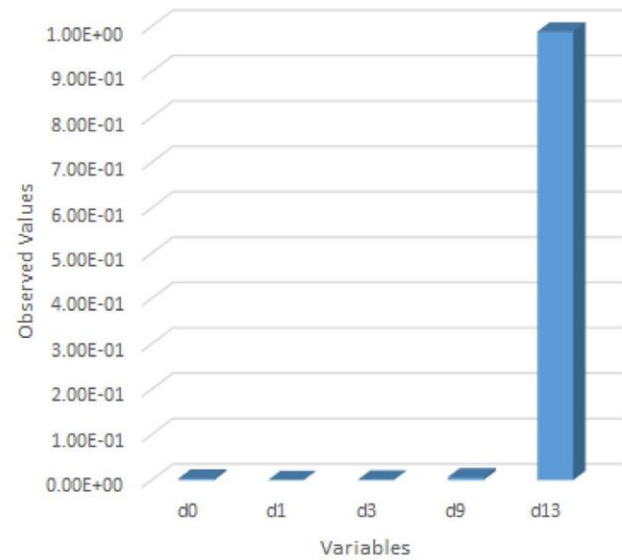
3.2. Modelo GEP para o Estado Indiano: Maharashtra

Maharashtra é o terceiro maior estado da Índia com uma área de 604,5 km² com uma população de mais de 110 milhões de pessoas. O estado sendo o segundo maior em termos de população tem várias pequenas áreas onde milhares de pessoas vivem em pequenos pedaços

de terra. Dharavi, a maior favela do mundo, também está localizada em Mumbai e atualmente é um hotspot COVID-19. Assim, torna-se necessário analisar o possível impacto do vírus em Maharashtra e prever como ele vai se comportar em um futuro próximo. Esta seção fornece detalhes sobre todos os três aspectos, incluindo resultados estatísticos para Maharashtra, Índia. Os resultados dos modelos de previsão em comparação com CC e DC reais em todo o estado são apresentados na Figura 6. Constatou-se que o modelo de previsão baseado em GEP proposto no presente trabalho fornece resultados muito confiáveis tanto para CC quanto para DC até 15 de maio de 2020. Os modelos gep prevêem ainda que até 25 de maio de 2020, o número total de CC chegará a quase 45,000 e no mesmo período, o dc total chegará a quase 1800. Assim, um aumento acentuado pode ser notado nos casos COVID-19 em Maharashtra nos próximos dias.

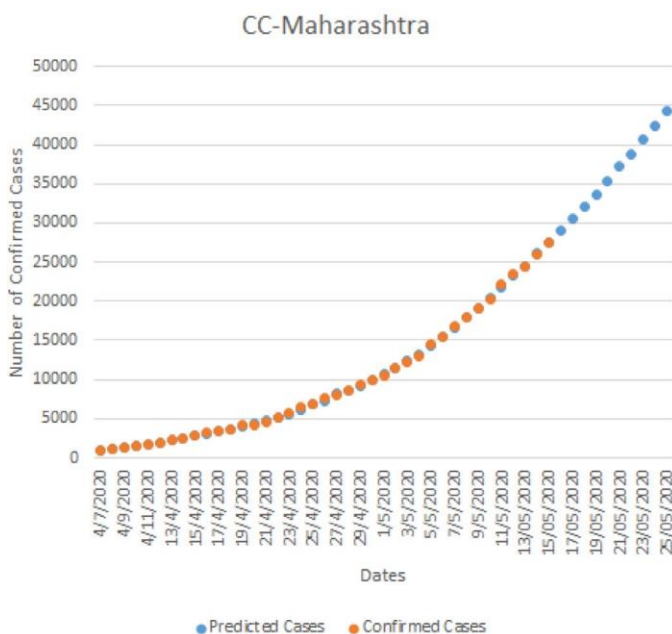


(a) Confirmed Cases

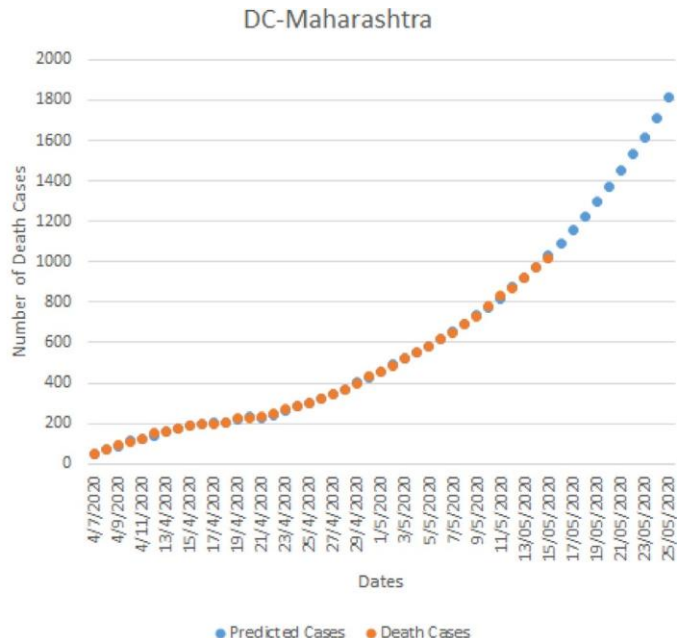


(b) Death Cases

Fig. 5. Contribuição das variáveis preditoras para COVID-19 na Índia.



(a) Confirmed Cases



(b) Death Cases

Fig. 6. Casos experimentais versus previstos para COVID-19 em Maharashtra usando modelo GEP.

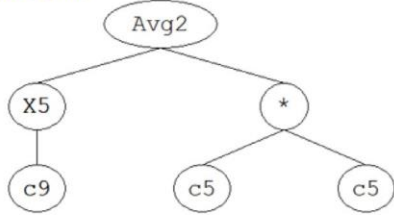
3.2.1. Validação baseada em árvore de expressão

ET é outro parâmetro importante e é útil na formulação matemática do problema em consideração. A partir da Figura 7, pode-se ver que os ETs para CC e DC em Maharashtra consistem em quatro sub-ETs independentes (genes ou subprogramas). Para CC, a função de vinculação utilizada é a função

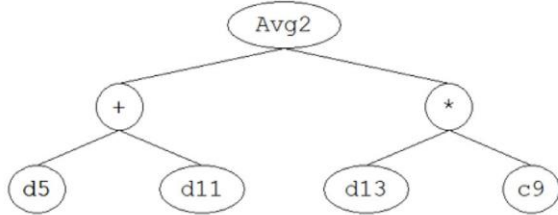
3.2.2. Importância Variável

Este parâmetro segue um fenômeno de randomização para cada um dos valores de entrada, a fim de analisar a importância de cada variável. Aqui a redução média em R^2 entre a produção prevista e desmedida é levada em consideração e a importância variável é calculada. Os resultados na Figura 8

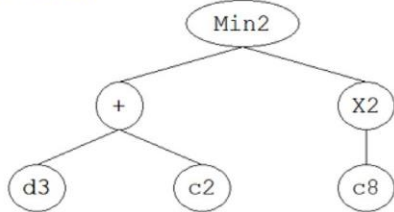
Sub-ET 1



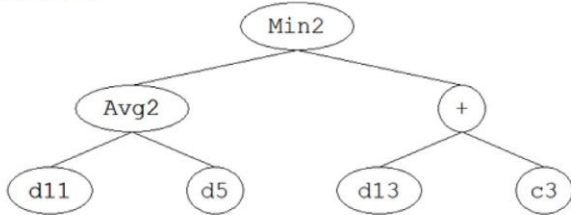
Sub-ET 2



Sub-ET 3

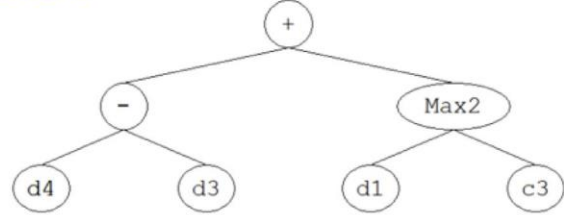


Sub-ET 4

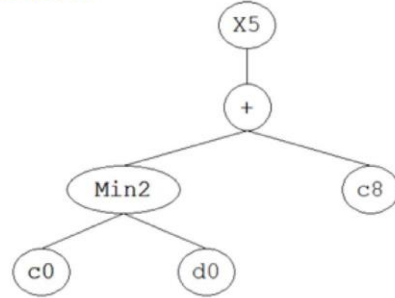


(a) ETs for Confirmed Cases

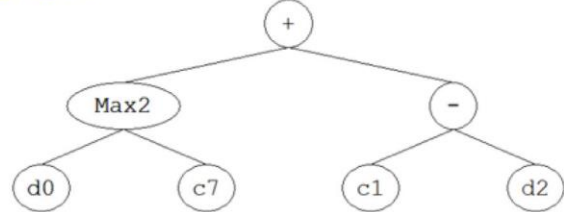
Sub-ET 1



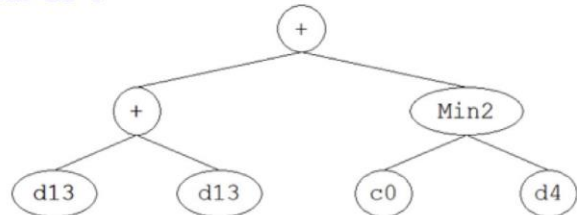
Sub-ET 2



Sub-ET 3



Sub-ET 4



(b) ETs for Death Cases

Fig. 7. Árvores de expressão (ETs) para COVID-19 em Maharashtra.

média em que, quanto à DC, a função de vinculação de adição é usada. Uma formulação generalizada dos ETs em termos de programa de simulação é apresentada no Algoritmo 3 para CC e Algoritmo 4 para DC.

mostram que a variável d_{13} desempenha um papel muito significativo e é o parâmetro mais importante para COVID-19 em Maharashtra para ambos os CC e DC. Além disso, nenhuma outra variável representa qualquer importância importante.

3.3. Modelo GEP para o Estado Indiano: Gujarat

Gujarat é o segundo estado mais afetado da Índia, depois que maharashtra e

para CC e DC. Os valores projetados indicam ainda que a possível disseminação do COVID-19 em Gujarat é um fator alarmante e precisa ser mantida em cheque. A validação baseada em ETs é apresentada na próxima

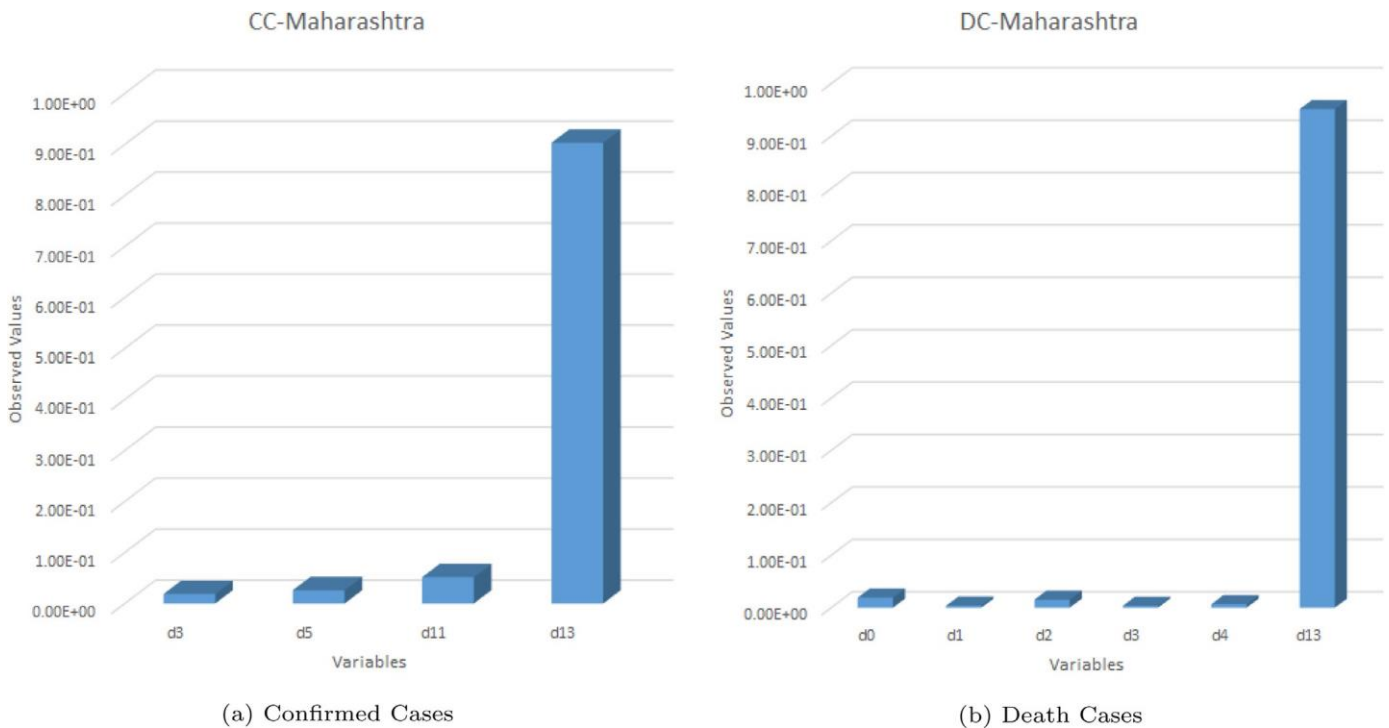


Fig. 8. Contribuição de variáveis preditoras para COVID-19 em Maharashtra.

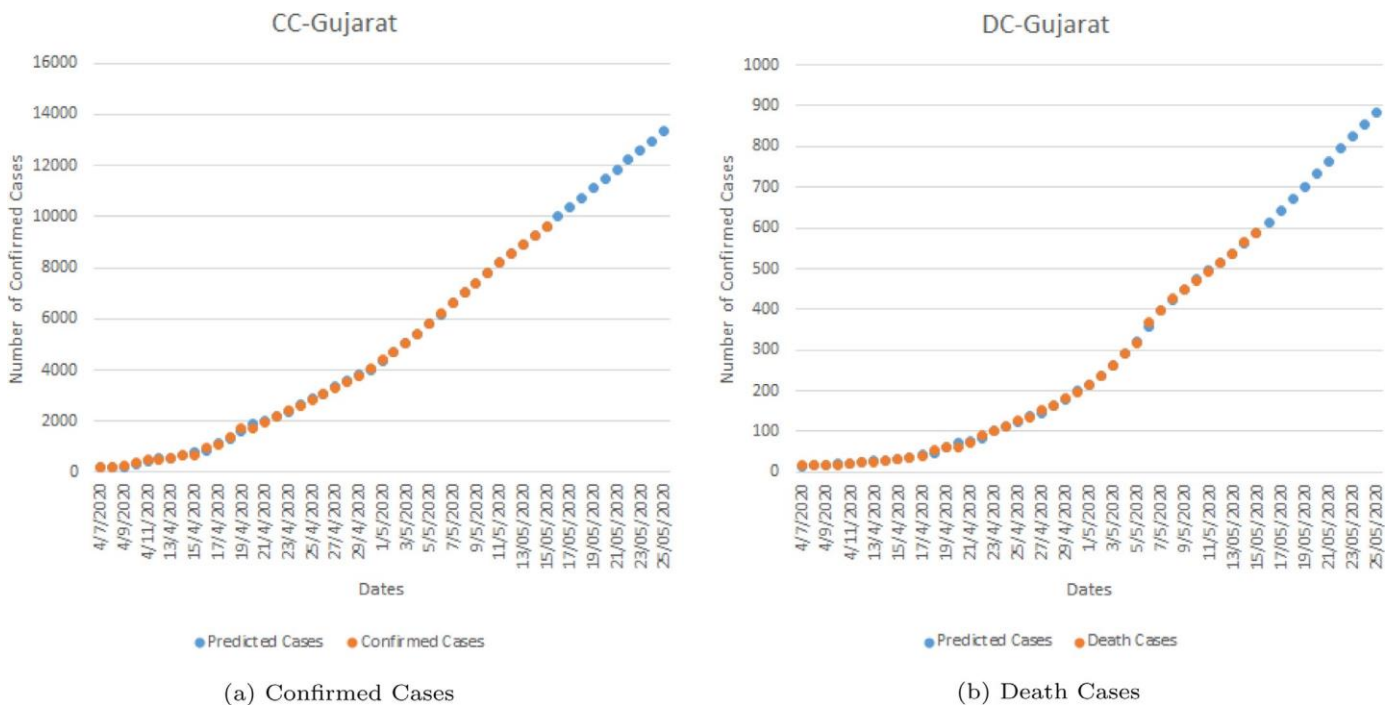


Fig. 9. Casos experimentais versus previstos para COVID-19 em Gujarat usando modelo GEP.

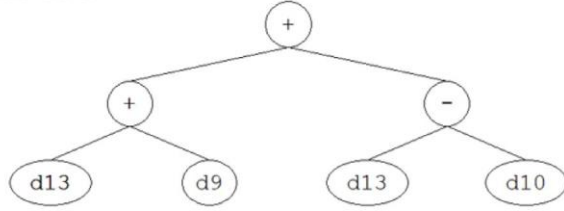
COVID-19 estão subindo a um ritmo mais alto. Esta seção propõe dois novos modelos GEP para CC e DC em Gujarat. Os resultados do modelo de previsão em relação aos casos reais tanto para CC quanto para DC são dados na Figura 9. A partir dos resultados, o número de CC para DC a partir de 15 de maio de 2020, é em torno de 10,0 0 a 600, respectivamente, e de acordo com os modelos de previsão gep, é expected para aumentar até 14,0 0 e 900, respectivamente,

subseção.

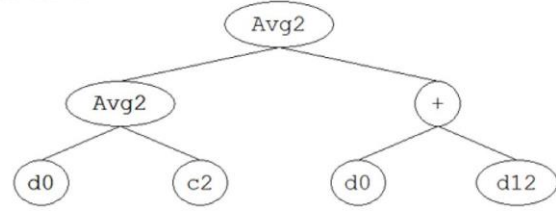
3.3.1. Validação baseada em árvore de expressão

Os ETs para CC e DC em Gujarat são dados pela Figura 10. Pode-se ver que os ETs para ambos os casos consistem em quatro subpro-

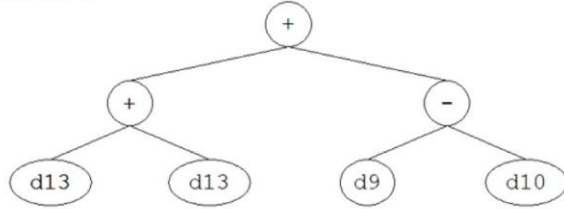
Sub-ET 1



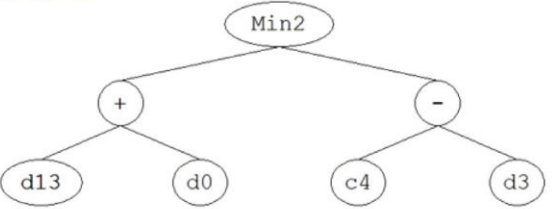
Sub-ET 1



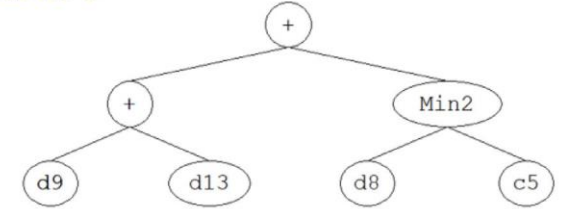
Sub-ET 2



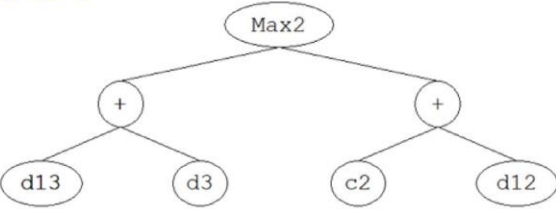
Sub-ET 2



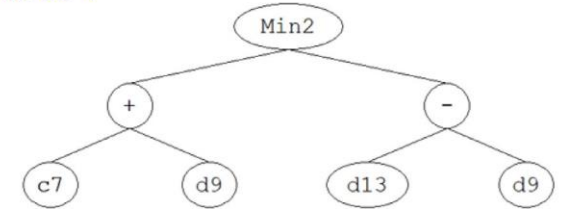
Sub-ET 3



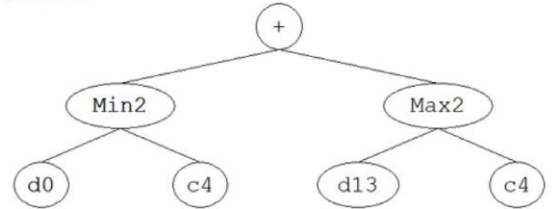
Sub-ET 3



Sub-ET 4



Sub-ET 4



(a) ETs for Confirmed Cases

(b) ETs for Death Cases

Fig. 10. Árvores de expressão (ETs) para COVID-19 em Gujarat.

Algoritmo 3 Modelo de previsão da Série Tempo gerado para CC em Maharashtra.

função Resultado=GEPModel(d) G1C9 = 6,15575133797683;
G1C5 = -14,2612512604032;
G2C9 = 7,58197580894619;
G3C2 = -3,60904377049443;
G3C8 = -99,5559674912138;
G4C3 = -949,636280409553; y = 0 .
0;

$y = (((G1C9^5) + (G1C5 * G1C5)) / 2 . 0); y = (y + (((d(6) + d(12)) + (d(14) * G2C9)) / 2 . 0)) / 2 . 0; y = (y + \min((d(4) + G3C2), (G3C8^2))) / 2 . 0; y = (y + \min(((d(12) + d(6)) / 2 . 0), (d(14) + G4C3)) / 2 . 0; Resultado=y; Fim$

gramas ou cromossomos ou simplesmente quatro sub-ETs. Todos os sub ETs são conectados por função de vinculação de adição para CC, enquanto para DC, a função de ligação de eragem av é usada. Além disso, a partir desses ETs, equações matemáticas podem ser formuladas de acordo com o requisito dos usuários finais e posterior avaliação em nível cromossômico. Os modelos gerais de previsão de séries temporais para Gujarat no caso de CC são dados no Algoritmo 5 e para DC é dado pelo Algoritmo 6. Aqui deve-se notar que o modelo foi gerado com base em 39 registros de treinamento tanto para CC quanto para DC, respectivamente.

3.3.2. Importância Variável

A partir da Figura 13, pode-se ver que as variáveis mais significativas no caso de Gujarat é d 13 tanto para CC quanto DC a partir de 15 de maio de 2020. Note-se que os resultados obtidos para o modelo de redução em termos de R^2 são normalizados de modo que sua adição faz a contagem 1. Além do d 13, para CC d 9 também desempenha um papel muito significativo, enquanto para DC nenhum outro parâmetro post qualquer significado. Assim, no geral, nós

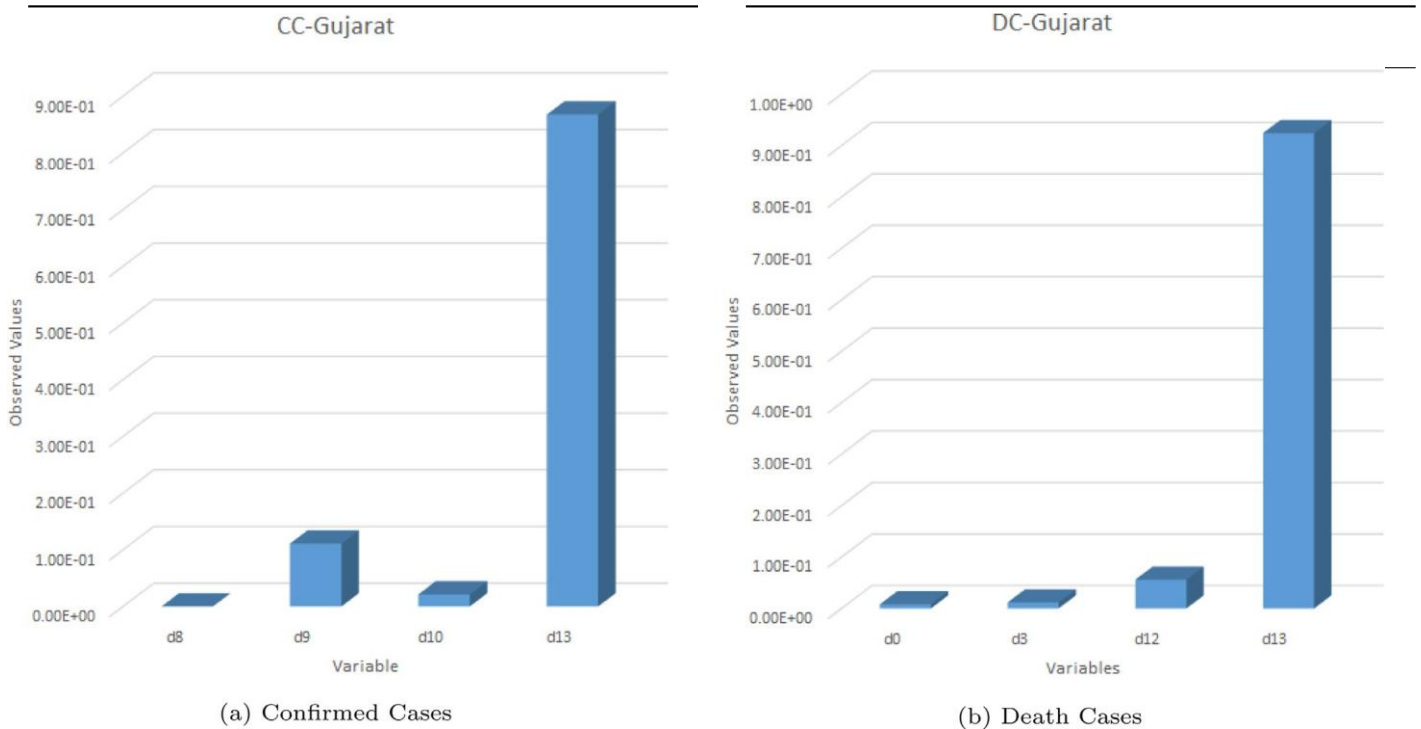


Fig. 11. Contribuição das variáveis predictoras para COVID-19 em Gujarat.

Algoritmo 4 Modelo de previsão da Série Hora gerado para DC em Maharashtra.

```

função Resultado=GEPModel(d) G1C3
= 32,1015827479207;
G2C8 = -0,285018566448923;
G2C0 = 2,9370445248558;
G3C7 = 28,7934042071809;
G4C0 = 6,4004324040421; y =
0. 0;

y = ((d(5) - d(4)) + máx (d(2) , G 1C 3)); y = (y + ((min
(G 2C 0 , d(1)) + G 2C 8) ^ 5) / 2. 0; y = (y + (máx (d(1) , G
3C 7) + (G 3C 1 - d(3))) / 2. 0; y = (y + ((d(14) + d(14)) +
min (G 4C 0 , d(5))) / 2. 0; Resultado=y;

```

Algoritmo 5 Modelo de previsão da Série Temporal gerado para CC em

Fim

Gujarat.

```

função Resultado=GEPModel(d) G3C5
= 377,923223116395;
G4C7 = 32,34584 874 91905; y
= 0. 0;

y = ((d(14) + d(10)) + (d(14) - d(11))); y = (y + ((d(14) +
d(14)) + (d(10) - d(11))) / 2. 0; y = (y + ((d(14) + d(14))
+ (d(10) - d(11))) / 2. 0; y = (y + min ((G 4C 7 + d (10))
, (d (14) - d(10))) / 2. 0;

```

Resultado=y;

Fim

pode-se dizer que o modelo GEP para Gujarat no caso de CC é sensível a duas variáveis d 13 e d 9, enquanto para DC apenas d 13 pose significativa cant. Na próxima subseção, o modelo GEP para Delhi é proposto.

Algoritmo 6 Modelo de previsão da Série Tempo gerado para DC inn Gujarat.

```

função Resultado=GEPModel(d) G1C2 = -
24,835755519817;
G2C4 = 536,083840303753;
G3C2 = 7,91741691335795;
G4C4 = 16,4523539142177; y = 0.
0;

y = (((d(1) + G 1C 2) / 2. 0) + (d(1) + d(13)) / 2. 0); y =
(y + min ((d(14) + d(1)), (G 2C 4 - d(4))) / 2. 0; y = (y +
max ((d(14) + d(4)), (G 3C 2 + d(13))) / 2. 0;

y = (y + (min (d(1) , G 4C 4) + máx (d(14) , G 4C 4)) / 2. 0;
Resultado=y;

```

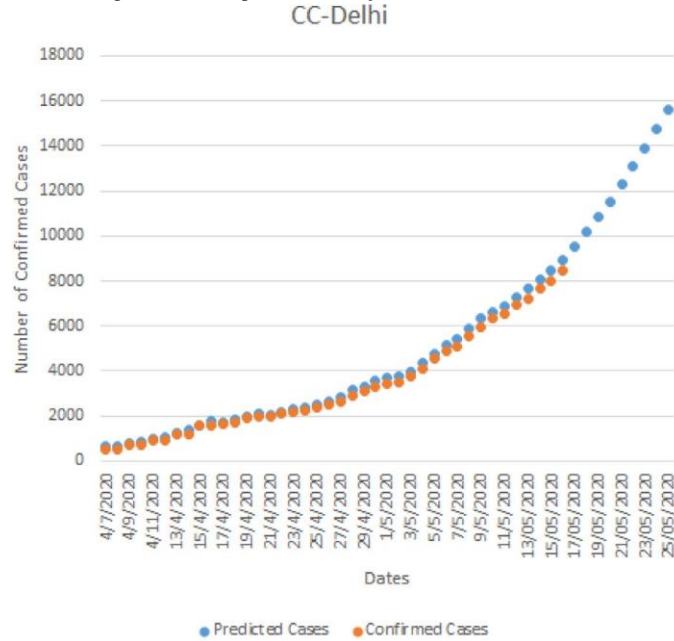
Fim

3.4. Modelo GEP para o Estado Indiano: Delhi

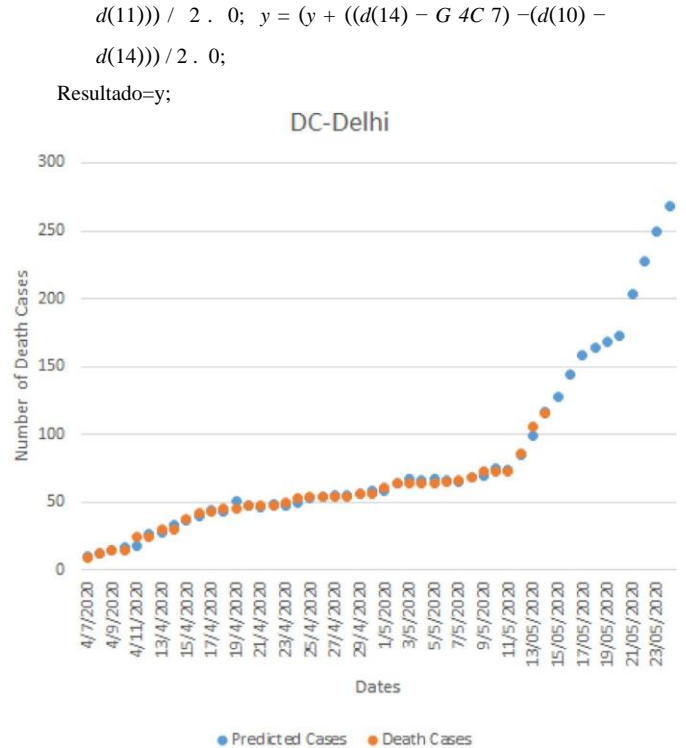
Délhi é uma das cidades mais populosas do país, com 1,48 km² e uma população de quase 20 milhões de pessoas. Assim, uma grande diversidade de população vive em um espaço de convivência tão compacto. E com o início do COVID-19, torna-se realmente muito importante analisar e prever como o vírus se comportará e se espalhará por todo o estado nos próximos dias. Com um total de mais de 8,0 0 CC e perto de 120 DC a partir de 15 de maio de 2020, um modelo como o GEP pode ser considerado como um importante fator na previsão...

ing a extensão do vírus. A partir dos resultados na Figura 12, pode-se ver que o vírus deve aumentar a uma taxa alarmante com um aumento total esperado no

CC para cerca de 16,0 0 0 e DC para cerca de 275 até 25 de maio de 2020. A curva está subindo exponencialmente e medidas necessárias precisam ser tomadas para estabilizar o mesmo. A validação baseada em ET para CC e DC em Delhi é apresentada na próxima subseção.



(a) Confirmed Cases



(b) Death Cases

Fig. 12. Casos experimentais versus previstos para COVID-19 em Delhi usando o modelo GEP.

3.4.1. Validação baseada em árvore de expressão

Os ETs no caso de Delhi também são agrupados em quatro sub-ETs. Aqui a função de vinculação de subtração é utilizada tanto para CC quanto para DC e é apresentada na Figura 13. Juntamente com a função de ligação de subtração, no caso de DC, a função de ligação exponencial também desempenha um papel significativo. Com base nesses ETs, as equações matemáticas podem ser formuladas e novas análises preditivas podem ser apresentadas. Os modelos de previsão de séries temporais, assim, gerados para CC e DC são dados pelo Algoritmo 7 e algoritmo 8, respectivamente. O vari...

Algoritmo 7 Modelo de previsão da Série Hora gerado para CC em Delhi.

função Resultado=GEPModel(d)

G1C6 = 524,898730692655; y = 0.

0;

y = (d(11) - más (d(10), G 1C 6));

y = (y + max ((d(11) - d (2)), ((d (8) + d(2)) / 2 . 0)) / 2 . 0; y =

(y + ((d(14) - d(8)) + (d(14) + d(14))) / 2 . 0; y = (y + max ((d(13)

- d (1)), ((d (6) + d(14)) / 2 . 0)) / 2 . 0;

Resultado=y

Fim

Algoritmo 8 Hora Série predição modeloggerado dur DC e Deli, o

função Resultado=GePModeld

G2C1 = 6.00024414807581;

G3C9 = -6,45508387551335;

G4C7 = -4,6340400982696; y = 0.

0;

y = exp(ceil(gep3 Rt(d(9)))) ; y = (y + ((d(8) - d(7)) - máx

(d(5), G 2C 1)) / 2 . 0; y = (y + (piso(d(1)) - (G 3C 9 -

$$d(11))) / 2 . 0; y = (y + ((d(14) - G 4C 7) - (d(10) - d(14))) / 2 . 0;$$

Resultado=y;

Fim

importância capaz de cada uma das variáveis de previsão é apresentada na próxima subseção.

3.4.2. Importância Variável

As variáveis preditoras no caso de Delhi desempenham um papel muito significativo. Aqui d 13 desempenha o papel mais significativo para ambos os modelos CC e DC GEP. Junto com isso, d 5 e d 12 também tem pouco impacto no modelo de previsão para CC, enquanto d 8, d 9 e d 10 representam pouco conhecimento significativo para o modelo DC. Os resultados são apresentados na Figura 14 e pode-se dizer que as variáveis de previsão para Delhi de duas a três variáveis de predição afetam os modelos GEP. Aqui também, os resultados do modelo de previsão são normalizados para que a adição de todas as variáveis equivale a 1. A próxima subseção de-tails sobre os resultados estatísticos para todos os casos em consideração.

3.5. Resultados Estatísticos para todos os casos

Os modelos calibrados nas seções acima utilizando ETs, modelagem GEP e parâmetros de importância variável só são aceitáveis se forem estatisticamente significativos. Como este é um dos primeiros estudos sobre o conjunto de dados COVID-19, não está disponível na literatura uma comparação com outras técnicas. Assim, no presente trabalho, um estudo comparativo para cc e DC para todos os três estados e toda a Índia é levado em consideração. Os resultados são apresentados em termos de RMSE e R. Já foi discutido nas seções previ-ous que a RMSE deve ser maior e R deve estar perto de 1. Os resultados para comparação são apresentados na Tabela 3. Aqui pode-se ver que para quase todos os casos, um RMSE mais alto foi encontrado e o valor de R para quase todos os casos é perto de 1. Os menores valores de R são de 0,9881, o que corresponde ao DC em Delhi, enquanto o maior valor é de 0,9999, o que corresponde a o CC em toda a Índia. Assim, a partir dos resultados estatísticos, é evidente que os modelos GEP propostos são altamente confiáveis e novas previsões podem ser derivadas com base

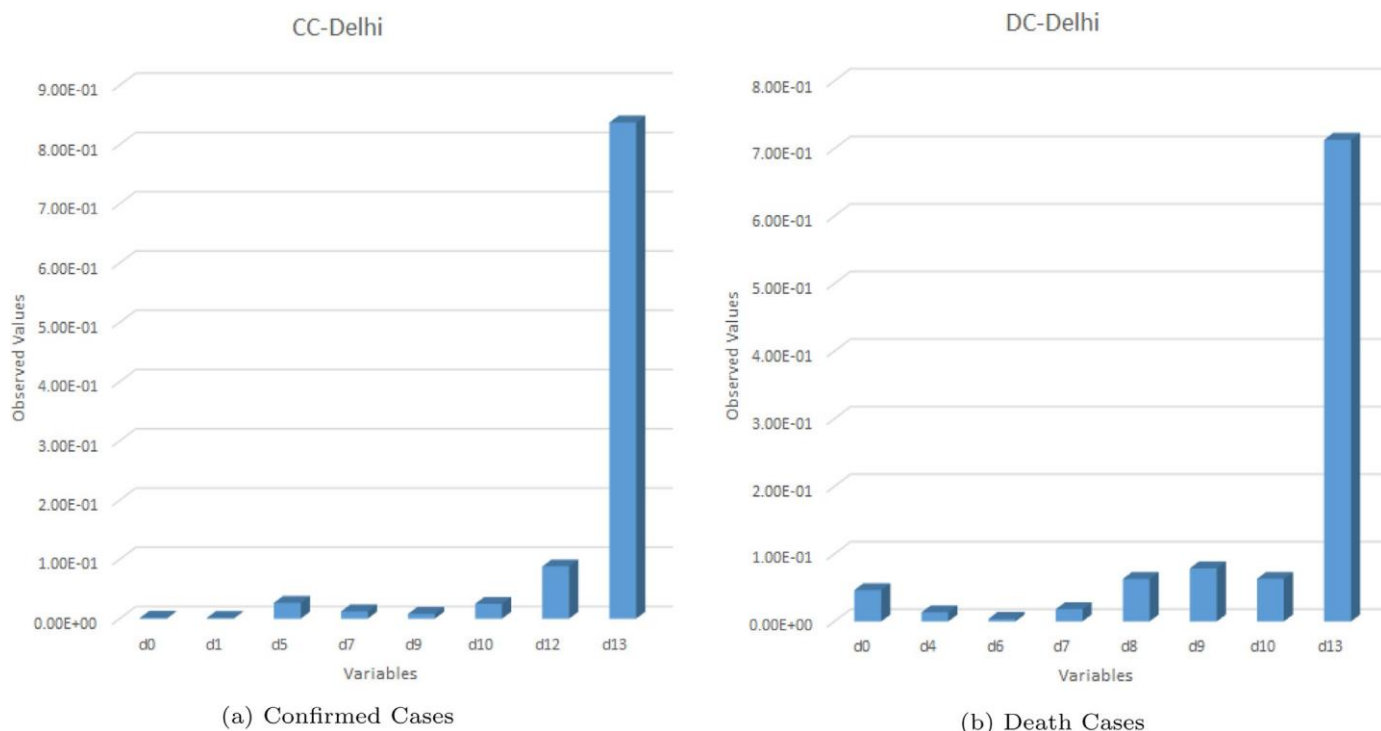


Fig. 14. Contribuição das variáveis predictoras para COVID-19 em Delhi.

4. Conclusão

Uma variante robusta e confiável do GEP foi usada para modelar os casos confirmados e casos de morte de COVID-19 na Índia. Novos modelos empíricos precisos foram projetados para previsão de CC e DC em toda a Índia e três grandes estados que são altamente afetados pela pandemia COVID-19. Estes estados incluem Maharashtra, Gujarat e Delhi. Os modelos propostos foram desenvolvidos a partir dos relatórios diários de situação dos casos COVID-19 publicados pelo Ministério do Interior, Govt. da Índia desde o início do primeiro bloqueio no país que é 24 de março de 2020. As seguintes conclusões foram formuladas com base nos modelos propostos:

- Os modelos GEP propostos neste trabalho são altamente confiáveis na previsão de casos confirmados e casos de morte em toda a Índia. Eles também satisfazem todos os requisitos de validação externa e, portanto, podem ser usados para prever casos futuros.
- Os valores *RMSE* e *R* para todos os casos são maiores e próximos a 1, respectivamente. Verificando assim a qualidade da solução dos modelos propostos e, portanto, aumentando as chances de previsões confiáveis.
- Os ETs derivados são equações matemáticas muito simples e básicas podem ser formuladas a partir deles sem qualquer implementação de laboratório demorada. Essas equações matemáticas podem ser ainda mais usadas para otimizar o modelo proposto usando técnicas de otimização difusora, como evolução diferencial, algoritmo de busca de cuco e Outros.
- As variáveis de previsão de todos os modelos propostos desempenham um papel muito significativo e descobriu-se que, além de Delhi, todos os outros modelos têm efeito de apenas um ou dois previsões variáveis. Tornando os modelos menos sensíveis às variáveis.
- Além disso, a partir dos resultados experimentais, pode-se dizer que os modelos GEP são altamente confiáveis, pois são baseados em dados experimentais, em vez de apenas suposições básicas, como no caso de modelos convencionais. Outra característica importante do modelagem GEP é que ele pode trabalhar em dados de séries menos temporais e ainda fornecer resultados confiáveis.

Assim, no geral, pode-se dizer que os modelos baseados em GEP são altamente re-responsáveis e podem ser tratados como referência para previsões de séries temporais. A preocupação surge quando o número total

de casos aumenta muitas vezes. Nesses casos, os modelos GEP precisam ser otimizados. Assim, equações matemáticas derivadas dos modelos GEP são otimizadas, otimizando algoritmos evolutivos de última geração altamente eficazes. Como direção futura, essas equações podem ser derivadas e algoritmos como algoritmo de rebanho Krill, algoritmo de rato-mole e outros podem ser usados para otimizar os modelos de previsão. Além disso, os modelos de pré-dição mostram CC e DC para os próximos 10 dias mostram que medidas rigorosas precisam ser tomadas para manter o vírus sob verificação. Aqui o bloqueio e o distanciamento social devem ser rigorosamente seguidos para que o vírus possa ser controlado e abster-se apenas para determinadas áreas.

Declaração de Interesse Concorrente

Os autores declaram que não têm interesses concorrentes ou relações pessoais que possam ter aparecido para influenciar o trabalho relatado neste artigo.

Referências

- [1] Huang C., Wang Y., Li X., Ren L., Zhao J., Hu Y., Zhang L., Fan G., Xu J., Gu X., et al. Características clínicas de pacientes infectados com o novo coronavírus 2019 em wuhan. *China Lancet* 2020;395:497-506.
- [2] QUEM.. Declaração sobre o agrupamento de casos de pneumonia em Wuhan, china;. Organização Mundial da Saúde: Genebra, Switzerland; 2020. Disponível online: <https://www.who.int/china/news/detail/09-01-2020-que-declaração-relativa-cluster-de-pneumonia-casos-in-wuhan-china> (acessado em 17 de fevereiro de 2020)
- [3] QUEM.. Novo coronavírus-Tailândia (ex-china). Organização Mundial da Saúde: Genebra, Suíça; 2020. Disponível online: <https://www.who.int/csr/don/14-janeiro-2020-novel-coronavirus-tailandia-ex-china/en> (acessado em 17 Fevereiro de 2020)
- [4] A abertura do diretor-geral da OMS é na coletiva de imprensa sobre o COVID-19 – 11 de março de 2020. 2020. [Online; acessado 21-março-2020].
- [5] Moore M., Gelfeld B., Okunogbe A.T., Christopher P.. Identificação de futuros pontos quentes de desafiabilidade: Índice de vulnerabilidade de doenças infecciosas; Corporação RAND: Santa Monica, CA, EUA. 2016. Disponível online: <https://www.rand.org/pubs/research-reports/RR1605.html> (acessado em 17 de fevereiro de 2020).
- [6] QUEM. Relatório de situação; Organização Mundial da Saúde: Genebra, switzer-land. 2020. Disponível online: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>.

- [7] Riou J, Althaus CL. Padrão de transmissão humana-humana precoce de wuhan 2019-ncov. *bioRxiv*; 2020.
- [8] Backer JA, Klinkenberg D, Wallinga J. O período de incubação de 2019-ncov em fezes entre viajantes de wuhan. *China medRxiv* 2020.
- [9] Lancet T. Índia sob bloqueio COVID-19. *Lancet* (Londres, Inglaterra) 2020;395(10233):1315.
- [10] Ministério do Interior. Relatório de situação: Governo da Índia. 2020. Disponível online: <https://www.mohfw.gov.in/>.
- [11] Chang S.L. Transmissão de modelagem e controle da pandemia COVID-19 na Austrália. 2020. *arXiv*:2003.10218.
- [12] Leia JM, Bridgen JRE, Cummings DAT, Ho A, Jewell CP. Novo coronavírus 2019-ncov: estimativa precoce de parâmetros epidemiológicos e pré-dições epidêmicas. *medRxiv* 2020.
- [13] Boldog P. Risco de assessment de novos surtos de coronavírus COVID-19 fora da China. *Journal of clinical medicine* 2020;9.2:571.
- [14] Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung S-M, Yuan B, Kinoshita R, Nishiura H. Características epidemiológicas da nova infecção por coronavírus: Uma análise estatística dos dados de casos disponíveis publicamente. *medRxiv* 2020.
- [15] Zheng Q., Meredith H., Grantz K., Bi Q., Jones F., Lauer S.. Equipe de IDD da JHU. estimativa em tempo real do novo tempo de incubação do coronavírus. 2020. Disponível na linha: <https://github.com/HopkinsIDD/ncov-incubation> (acessado em 17 de fevereiro de 2020).
- [16] Eubank S., Guclu H., Kumar V.A., Marathe M.V., Srinivasan A., Toroczkai Z., Wang N.. Modelando surtos de doenças em redes sociais urbanas realistas. *Natureza*, 429, 6988, 180-184.
- [17] Liu T, Hu J, Kang M, Lin L, Zhong H, Xiao J, Deng A. Dinâmica de transmissão do novo coronavírus 2019. 2019-nCoV 2020.
- [18] Koza JR. Programação genética: Na programação de computadores por meio de seleção natural. Cambridge, MA: MIT Press; 1992.
- [19] Salgotra R, Singh S, Singh U, Saha S, Gandomi AH. COVID-19: Time series conjuntos de dados Índia versus mundo. *Mendeley Data* 2020;v1. doi: 10.17632/tmrs92j7pv.1.
- [20] Goldberg D.E., Holland J.H.. Algoritmos genéticos e aprendizado de máquina. 1988.
- [21] Banzhaf W, Nordin P, Keller R, Francone F. Programação genética é uma introdução. Sobre a evolução automática dos programas de computador e sua aplicação. Heidelberg, Alemanha/São Francisco: dpunkt/Morgan Kaufmann; 1998.
- [22] Ferreira C. Programação de expressão genética: Um novo algoritmo adaptativo para problemas de solv-ing. *Complexo Syst* 2001;13(2):87-129.
- [23] Fanelli D, Piazza F. Análise e previsão de expansão do COVID-19 na China. *Itália e França Caos, Solitons & Fractals* 2020;134:109761.
- [24] Mandal M, Jana S, Nandi SK, Khatua A, Adak S, Kar TK. Um estudo baseado em modelo na dinâmica do COVID-19: Previsão e controle. *Caos, Solitons & Fractals* 2020;109889.
- [25] Gandomi AH, Alavi AH, Mirzahassemi MR, Nejad FM. Modelos de base genética não linear para previsão do número de fluxo de misturas asfálticas. *J Mater Civ Eng* 2011;23(3):248-63.
- [26] Javadi AA, Rezaei M. Aplicações de inteligência artificial e técnicas de mineração de dados na modelagem do solo. *Geomech Eng* 2009;1(1):53-74.
- [27] Fair KM, Zachreson C, Prokopenko M. Criando uma rede de transporte substituto a partir de dados censitários do Bureau of Statistics australiano. *Dados científicos* 2019;6(1):1-14.
- [28] Pal R., Sekh A. Kar S., Prasad D.K.. Previsão de risco de país baseado em rede neural de COVID-19. 2020. *arXiv*:20 04.0 0959.
- [29] Gandomi AH, Babanajad SK, Alavi AH, Farnam Y. Nova abordagem para modelagem de comprimento de concreto sob compressão triaxial. *Revista de materiais em engenharia civil* 2012;24(9):1132-43.
- [30] Alavi AH, Gandomi AH. Uma abordagem robusta de mineração de dados para formulação de sistemas de engenharia geotécnica. *Eng Comput* 2011;28(3-4):242-74.
- [31] Ferramentas **GeneXpro**. Software de computador. Bristol, Reino Unido: GEPSOFT Ltd; 2006.
- [32] Smith GN. Probabilidade e estatística em engenharia civil. Collins, Londres 1986.
- [33] Frank Ie, Todeschini R. O manual de análise de dados. Amsterdã: Elsevier; 1994.
- [34] Golbraikh A, Tropsha A. Cuidado com o 2º trimestre! *J Mol Graphics Modell* 2002;20(4):269-76.
- [35] Roy PP, Roy K. "em alguns aspectos da seleção variável para modelos parciais de regressão de quadrados. *QSAR Pent Sci* 2008;27(3):302-13.
- [36] Gandomi AH, Alavi AH, Ryan C. Manual de programação genética aplica-a-tions. Cham: Springer; 2015.
- [37] Gandomi AH, Alavi AH. Rebanho krill: uma nova otimização bio-inspirada al-gorithm. *Comunicações em ciência não linear e simulação numérica* 2012;17(12):4831-45.
- [38] Salgotra R, Singh U. O algoritmo de rato-toupeira nu. *Computação Neural e Aplications* 2019;31(12):8837-57.