ISSN: •2708-6437(Online) •2708-6429(Print)

Journal of Engineering Advancements

Editor-in-Chief:
Prof. Dr. Mohammad Mashud

Volume 04 Issue 02

SciEn

Published by:
SciEn Publishing Group

Journal of Engineering Advancements
Apt. # 6 C-D, House # 191
Road # 12/A, Dhanmondi R/A
Dhaka-1209, Bangladesh
Email: jea@scienpg.com Website: www.scienpg.com/jea/

SciEn
Published in: June 2023
Published by: SciEn Publishing Group
Price: Each Issue BDT 200.00 (US$ 15) **ISSN:** 2708-6437 (Online) 2708-6429 (Print)

Journal of Engineering Advancements
Apt. # 6 C-D, House # 191
Road # 12/A, Dhanmondi R/A
Dhaka-1209, Bangladesh
Email: jea@scienpg.com Website: www.scienpg.com/jea/

Journal of Engineering Advancements

Volume 04, Issue 02 June 2023

CONTENTS

Original Articles

Journal of Engineering Advancements

Published by: SciEn Publishing Group

Apt. # 6 C-D, House # 191, Road # 12/A
Dhanmondi, Dhaka-1209, Bangladesh
Email Address: jea@scienpg.com

www.scienpg.com/jea/

This page is left intentionally blank

SciEn
Journal of Engineering Advancements Vol. 04(02) 2023, pp 35-39 https://doi.org/10.38032/jea.2023.02.001

Fault Diagnosis in Gas Lift System Using PDF Data

Ojonugwa Adukwu

Department of Industrial and Production Engineering, School of Engineering and Engineering Technology, Federal University of Technology, P.M.B.704, Akure, Ondo State, Nigeria

Received: June 03, 2023, Revised: June 13, 2023, Accepted: June 14, 2023, Available Online: June 25, 2023

ABSTRACT

Fault detection and isolation in the gas lift system were implemented assuming the gas lift variables are stochastic. Injection valve coefficient (C_{iv}), production choke coefficient (C_{pc}), annulus pressure (P_a), and wellhead pressure (P_{wh}) were observed to show variations with faults presence. By simulating these gas lift variables as stochastic, the probability density function (PDF) data were used to generate decision functions for both the detection and isolation of the gas lift valve faults. The scheme accurately detected and isolated faults in the injection valve coefficient (C_{iv}) and production choke coefficient (C_{pc}). The result of this diagnosis will aid the proper implementation of fault tolerant control in the gas lift system which will lead to its optimal operation.

Keywords: Gas Lift, Fault Detection, Fault Diagnosis, Fault Isolation, PDF Data.

1 Introduction

Gas lift systems are used for lifting crude oil into the production platforms when the reservoir pressure becomes insufficient [1]-[3]. Due to the location and the materials transported, gas lift usually suffers from faults that reduce its ability to function as desired or even cause danger. To optimally operate the gas lift system, therefore, a proper fault diagnosis procedure must be implemented.

Fig. 1 is a gas lift system. The key variables are the pressures, flow rates, and masses. The pressures are annulus pressure (P_a), reservoir pressure (P_r), bottomhole pressure (P_{bh}), well pressure (P_w), wellhead pressure (P_{wh}) and separator pressure (P_s). The flow rates are flow of lift gas from the compressor into the annulus (w_{gl}), flow from annulus into tubing (w_{iv}), flow from reservoir into the tubing (w_r) and flow from the tubing into the separator (w_{pc}). The masses are the mass of gas in the annulus (m_{ga}), mass of gas in tubing (m_{gt}) and mass of oil in tubing (m_{ot}). These masses form the states of the gas lift system.

The natural source of energy for the lift is the reservoir pressure (P_r) [4]-[5]. During the operation of the system, this pressure could become insufficient to lift the crude as desired. Gas from the compressor station is therefore supplied to the annulus through the gas lift valve. The gas then flows to the tubing through the injection valve and mixes with the liquid in the tubing. This lightens the tubing liquid ensuring adequate oil production again [6]-[7].

The valves control the flow of oil through the system [8]-[9]. The opening and closing of the valves expose the valves to faults that affect their optimal operation. Due to the variation of reservoir parameters in particular, the gas/oil ratio (GOR), the gas lift variables are stochastic in nature, and fault diagnosis using the variables is better implemented using probability density function (PDF) data similar to results in DC motor in [10].

In this article, faults in gas lift valves are detected and isolated. Variables that change in response to faults presence are selected. The residuals which follow Gaussian distribution hence permit the use of the PDF data to provide the decision functions presented. The decision functions to detect the fault are monitored. By generating fault signatures, the faults are isolated.

The remainder of this article is as follows: Section 2 discusses the materials and methods; Section 3 presents the results and discussion, and Section 4 concludes the article.

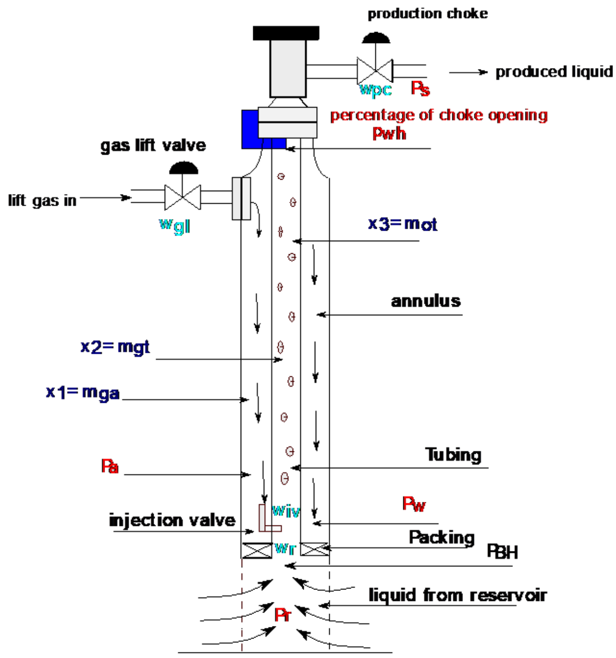

Fig. 1 A single-well gas lift system

2 Materials and Methods

2.1 Gas Lift faults

Fault diagnosis involves three activities namely: (i) detection, (ii) isolation, and (iii) identification. While fault detection implies detecting fault presence, fault isolation involves determining faulty components and fault isolation deals

*Corresponding Author Email Address: oaualaku@yahoo.co.uk
Published by: SciEn Publishing Group

with identification of fault type. Some valve faults can have a step-change effect on the flow rate. The valve coefficient is obtained by lumping together all the constants that affect flow rates through the valve. For a fluid of constant density at a given temperature, C_{iv} is the flow rate per unit pressure change (ΔP) if the percentage valve opening is constant. Hence, the flow rate through the valve is determined by the valve coefficients as well as the percentage of valve opening if the pressure, density, and temperature are fixed. Faults have effects on the control of flow rate using the valve coefficients as parameters.

Fault diagnosis is therefore implemented here by assuming valve faults to result in a step-like change in these valve coefficients. The other assumption is that the gas lift variables are Gaussian implying zero mean with nonzero variance.

Augmented states are therefore obtained if the valve coefficients are added to the masses described in section 1 as:

$$x = [m_{ga} \quad m_{gt} \quad m_{ot} \quad C_{iv} \quad C_{pc}]^T \qquad (1)$$

These augmented states vary for m_{ga}, m_{gt}, m_{ot} but remains constant for C_{iv}, C_{pc} until the arrival of faults.

2.2 Gas Lift Variables Used for Fault Detection

Fault detection involves generating a residual, r. This residual is constant at a steady state at each time sample in the absence of fault. This is still true even in the presence of input change and disturbance for an ideal system. This residual changes only when there is a fault. For a stochastic system, with Gaussian noise distribution, the residual has zero mean and nonzero variance. At no-fault case. The mean is however nonzero at all or some points when there is fault.

C_{iv} and C_{pc} are good candidates to be used to generate fault signatures. Other variables also vary in known ways due to faults in the valves. Fig. 2 shows an abrupt fault of 20% step increase in C_{iv} introduced at the 60th minute and C_{pc} introduced at the 120th minute while Fig. 3 shows the step decrease.

In Fig. 2 and Fig. 3 a step fault in C_{iv} has no affect C_{pc}, similarly, a step fault in C_{pc} does not affect C_{iv}. Therefore, if a residual is generated for a fault in the C_{iv}, it will not affect the residual for the fault in the C_{pc}. To improve detection and isolation, we increase the number of variables that will be used to generate fault signatures. We make an assumption to enable us to use this: there is no disturbance input and the system is operating at a steady state where the control input is constant throughout. Removing the disturbance effect helps avoid having to solve the optimization problem to decouple the disturbance effect hence easing our fault detection here. We, therefore, add other variables to the two gas lift valve coefficients. These additional variables are discussed briefly below:

P_a: The pressure in the annulus is directly affected by the injection valve coefficient (C_{iv}). All other things being constant, P_a decreases if C_{iv} increases owing to more gas flow from the annulus into the tubing which decreases the mass of gas in the annulus m_{ga}. But P_a is affected by other variables namely, the pressure in the well at the injection point (P_w) and the rate of gas injection into the annulus (w_{gl}). Despite these a change in C_{iv} will show symptoms in P_a hence it is selected as one of the variables for detection to a large extent. But since annulus pressure is also affected by the well pressure at the injection point, P_a is also affected by w_{pc} since it affects the P_w.

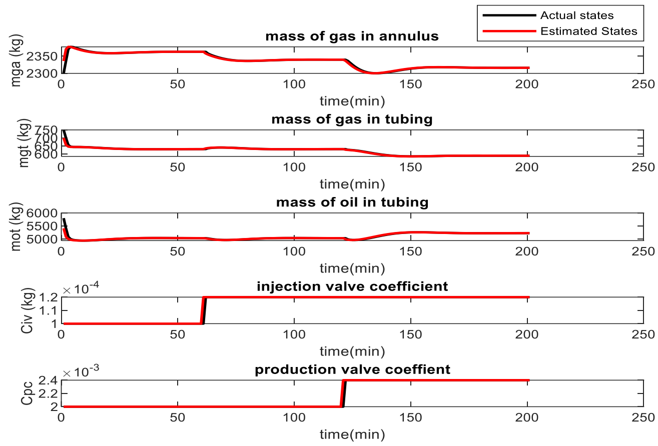

Fig. 2 Augmented gas lift states with a 20 % step increase in valve coefficients

Consequently, a fault in C_{iv} and C_{pc} show symptoms in P_a if a disturbance is removed and the system operates at constant input. This is shown where in both cases, the residual responded in the same way to fault presence, hence this is used for fault detection.

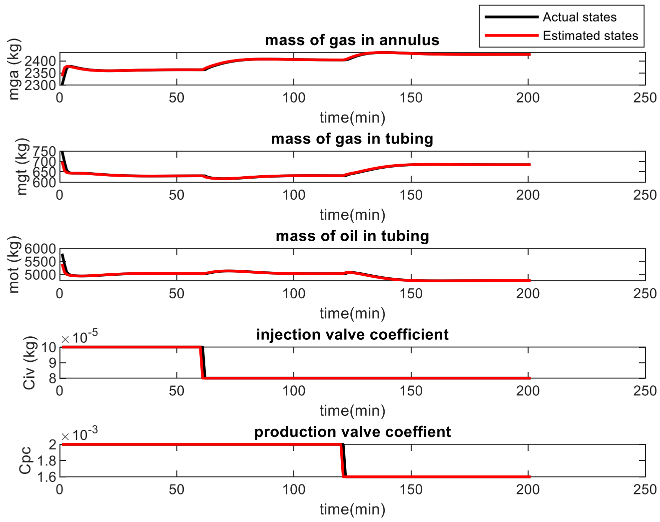

Fig. 3 Augmented gas lift states with a 20 % step decrease in valve coefficients

P_{wh}: The pressure of the wellhead is affected by the flow rate through the production choke w_{pc}. The pressure decreases as the production rate increases. P_{wh} depends on the mass of tubing gas which is contributed by flow from the reservoir and from the annulus. P_{wh} is hence affected by C_{iv} too but minimally as the loss in mass due to reduced w_{iv} can be compensated for by the flow from the reservoir. A fault in C_{iv} shows little effect on the P_{wh} as seen in Fig. 4(b) where the fault in C_{iv} affected the residual only during the fault transient period. A fault in C_{pc} shows strong effect on P_{wh} as shown in Fig. 5(b). P_{wh} can therefore be used for both detection and isolation when appropriate threshhold is selected.

Flowrates: Many other variables vary in an observable way with valve coefficients faults when the disturbance is ignored and input kept constant; for example, Fig. 6 shows the flow rates for the gases.

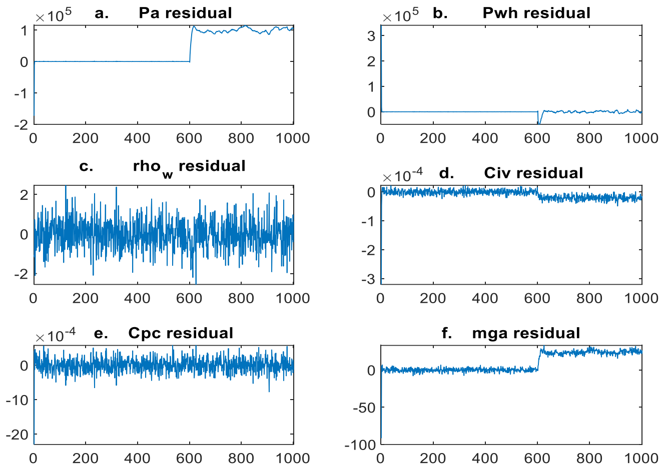

Fig. 4 Gas lift residual with a 20 % step increase in C_{iv}

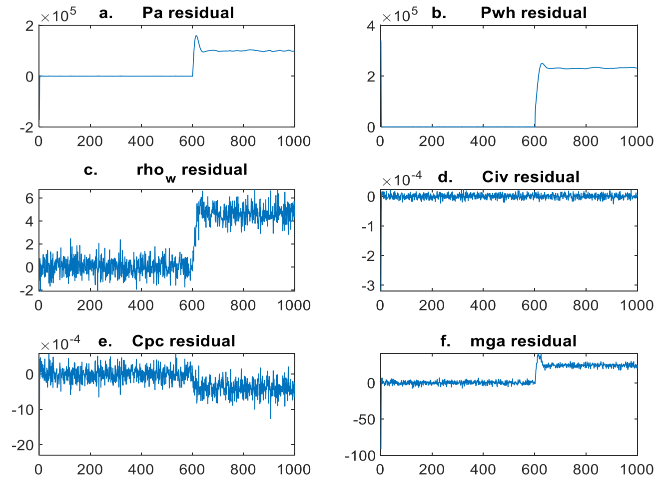

Fig. 5 Gas lift residual with a 20 % step increase in C_{pc}

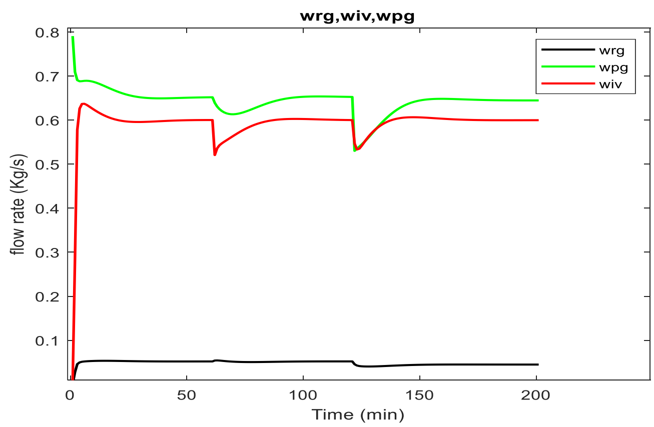

Fig. 6 Gas lift residual with a 20 % step increase in C_{pc}

Both w_{rg} and w_{pg} change with change in the valve coefficients at steady states while w_{iv} change only during the transient period. Flow through the injection valve is not affected at a steady state due to the fact that there is no net change in annulus gas implying that $w_{iv} = w_{gl}$ at steady state. The mass of gas in the annulus is also seen to vary with changes in C_{iv} as seen in Fig. 4.

The mass of gas in the annulus behaves similarly to P_a (Fig. 4(f) and Fig. 5(f)), hence no need to include it. It is expected that mixture density should respond to both a change in C_{iv} and C_{pc} but Fig. 4(c) shows no residual change with fault in C_{iv} while shows change with fault in C_{pc}. Fig. 4(d),(e) and Fig. 5(d),(e) all behave as expected, with each residual responding to change in the corresponding fault and unaffected by fault in other coefficients. We therefore limit the additional variables to two which are P_{wh} and P_a.

2.3 Gas Lift models.

The gas lift models presented here are adapted from [11]. The mass (differential equations):

$$\frac{dm_{ga}}{dt} = w_{gl} - w_{iv} \tag{2}$$

$$\frac{dm_{gt}}{dt} = w_{iv} + w_{rg} - w_{pg} \tag{3}$$

$$\frac{dm_{ot}}{dt} = w_{ro} - w_{po} \tag{4}$$

Flow rate:

$$w_{iv}=C_{iv}\sqrt{max(0,\ \rho_a(P_a\text{-}P_w)} \tag{5}$$

$$w_{pc}=C_{pc}\sqrt{max(0,\ \rho_w(P_{wh}\text{-}P_s)}\ f(u) \tag{6}$$

$$w_{pg}=\frac{\lambda}{\lambda+1}w_{pc} \tag{7}$$

$$w_{po}=\frac{1}{\lambda+1}w_{pc} \tag{8}$$

$$w_{ro}=C_r\sqrt{\rho_o(P_r\text{-}P_{bh})} \tag{9}$$

$$w_{rg}=GOR \tag{10}$$

The pressure:

$$P_a=(\frac{T_aR}{V_aM}+\frac{g}{A_a})ga \tag{11}$$

$$P_{wh}=\frac{T_tR}{M}\left(\frac{m_{gt}}{V_t+V_{bh}\text{-}\frac{m_{ot}}{\rho_l}}\right) \tag{12}$$

$$P_w=P_{wh}+\frac{\left(m_{gt}+m_{ot}\text{-}\rho_mV_{bh}\right)gH_t}{V_t} \tag{13}$$

$$P_{bh}=P_w+\rho_m gH_{bh} \tag{14}$$

The density:

$$\rho_a=\frac{m_{ga}P_a}{T_aR} \tag{15}$$

$$\rho_m = \frac{m_{gt} + m_{ot} - \rho_o L_{bh} A_{bh}}{L_w A_w} \tag{16}$$

3 Results and Discussion

3.1 Residual Generation

The residual has been shown in Fig. 4 and Fig. 5 of section 2.2 to change in reaction to a fault. Fig. 7 shows the decision function for the residuals for a no-fault case. It is seen that at no fault, the function variation is very small hence no fault can be detected. When a fault is introduced, these functions change in response to the faults as shown later in section 3.3.

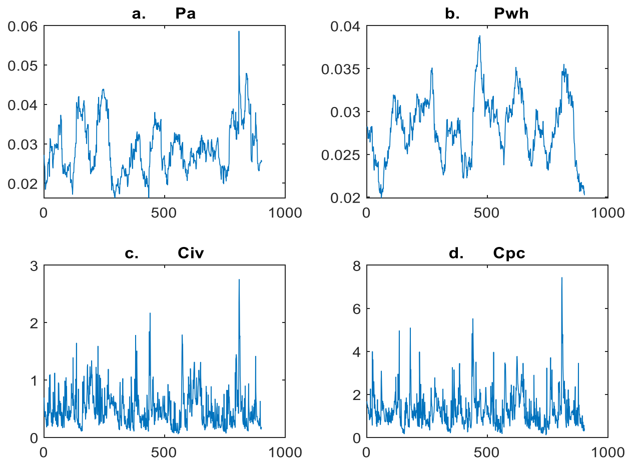

Fig. 7 No fault decision function based on the residuals

3.2 Hypothesis test using PDF data

ResidualS statistics of the gas lift system change following the arrival of fault. Fig. 8 is the PDF plot of the residual for the selected variables. The mean and standard deviation of some of the variables change with a fault when the production choke valve is increased by 20 % after the 600th minute. While for annulus pressure and wellhead pressure, the mean of the PDF shifts right following the fault, the PDF of the residual for the production choke characteristics shifts left after the fault, and that of the injection valve remains constant. The variation of the PDF of the residuals of some variables of the gas-lifted system with fault present is used to produce a decision function. This decision function is monitored which when an appropriate threshold is exceeded, a fault is detected, and the alarm rings.

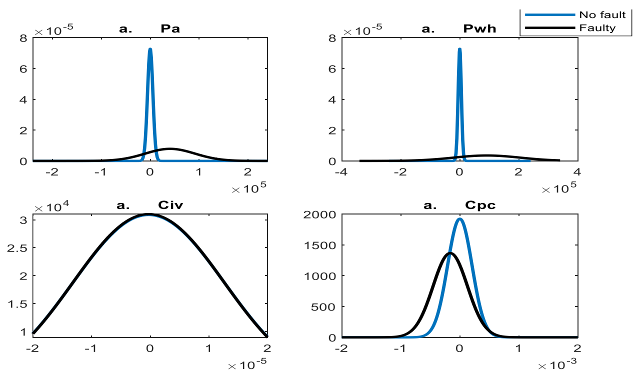

Fig. 8 PDF of the residuals for faulty and no-fault cases

The hypotheses according to [12], this hypothesis test is stated as follows:

H_0: $\Theta(i) = \Theta_0$ for $1 \leq i \leq t$

H_1: $\Theta(i) = \Theta_0$ for $1 \leq i \leq i_f$

$\Theta(i) = \Theta_1$ for $i_f \leq i \leq t$

Where i is the time instant, t is the simulation time and i_f is the fault occurrence time. We select H_0 if there is no change in decision function exceeding the selected threshold over the whole simulation time, else, H_1 is selected.

3.3 Fault detection using PDF data

The decision function for the fault of a 20 % increase in C_{iv} occurring at the 600th sample time is given in Fig. 9 while a 20 % increase in C_{pc} is given in Fig. 10. The data window is taken as N = 50 samples (minutes) and the simulation time is 1000 samples with the first 100 minutes removed to avoid the transient part (caused by the initial condition). In Fig. 9, only the annulus pressure P_a and C_{iv} decision function change in response to a fault. But in Fig. 10 the decision function corresponding to the annulus pressure, wellhead pressure and C_{pc} responded to the faults while the decision function for the C_{iv} does not respond to the C_{pc} fault.

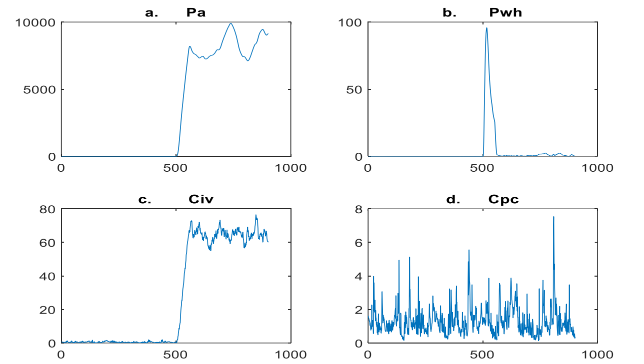

Fig. 9 Decision function for a fault of 20% in the C_{iv}

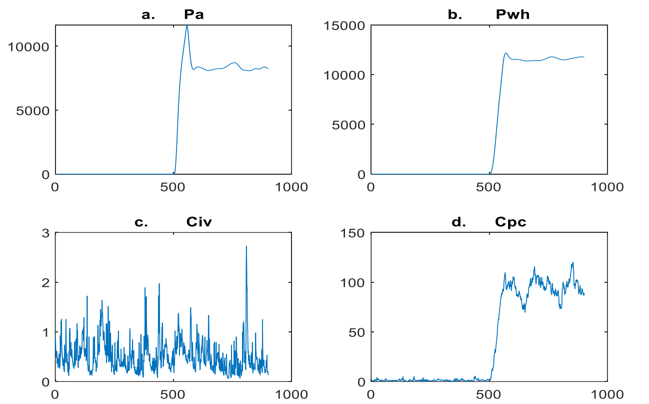

Fig. 10 Decision function for a fault of 20% in the C_{pc}

This makes it much easier to isolate faults in C_{pc} then it is for C_{iv}. Fig. 9 and Fig. 10 show that a threshold of h = 23 selected can easily be met with little detection delay. Fig. 11 and Fig. 12 show the zoomed sensitive residuals to indicate the point at which the fault is detected (time at which gk = h = 23) for Fig. 9 and Fig. 10 respectively, where gk implies decision function.

In Fig. 11 the gk for the P_a residual crossed h = 23 at 504 samples which is the detection time implying that the detection delay is therefore 3 samples, while the detection delay for the C_{iv} is 22 samples. Similarly, the detection delays in Fig. 12 are 3, 3, and 19 samples for P_a, P_{wh} and C_{pc} respectively. All these detection delays are within acceptable values for a gas lift system.

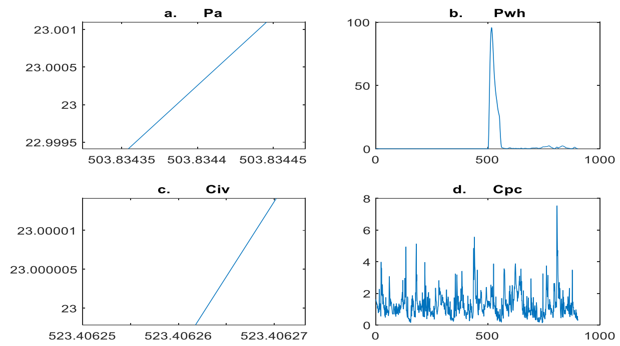

Fig. 11 Zoomed decision function for a fault of 20% in the C_{iv}

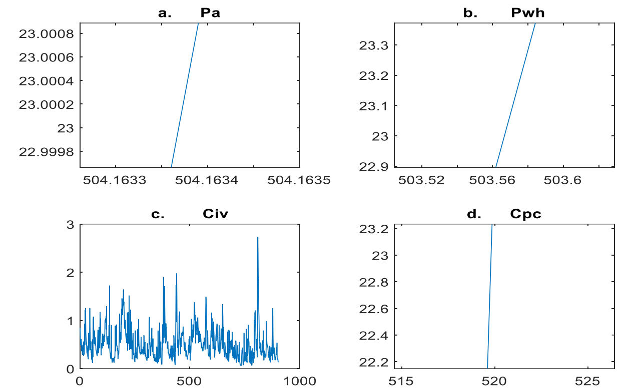

Fig. 12 Zoomed decision function for a fault of 20% in the C_{pc}

The estimated change magnitude of the parameter is 2.1*E-4 for C_{pc} and 5.7*E-5 for C_{iv}. If we note that C_{pc} provided is 2*E-3 and C_{iv} is 1E-4, then the estimated changed magnitude of the fault is 10 % for C_{pc} and 57 % for C_{iv}. But we apply a 20% change in both cases implying the effects of noise and the size of N has affected the estimated parameter at the point the fault is detected.

3.4 Fault isolation in gas lift system using PDF data.

For fault isolation, Table 1 shows the fault signatures for the variables and parameters selected. A 1 in the entry for a given residual implies the decision function changes following the arrival of the fault and 0 means the decision function is insensitive to a fault. From Table 1, each fault can be easily isolated since they have different signatures. The residual for the annulus pressure only helps in detection but not isolation as it is sensitive to both faults.

Table 1 Fault signatures for the gas lift system

	fault in C_{iv}	fault in C_{pc}
gk for P_a residual	1	1
gk for P_{wh} residual	0	1
gk for C_{iv} residual	1	0
gk for C_{pc} residual	0	1

The decision presented in Fig. 9 to Fig. 12 evaluates the residual and a simple logic is used to isolate the faults based on Table 1. The residual for P_a can be ignored as it is sensitive to both faults and a necessary condition for isolation is that the number of residuals must be at least equal to the number of faults. Even if P_a is removed, there are still three residuals for two faults.

4 Conclusion

Fault detection and isolation in the gas lift system was implemented considering the selected variables as stochastic hence the PDF data of these variables were used to generate fault signatures. The PDF data of the variables were used to generate decision functions used for both fault detection and isolation. The valve coefficients were used in addition to annulus pressure, flow through the injection valve, and pressure of the wellhead. The algorithm detected and isolated all fault scenarios examined for the gas lift system. The detection and isolation of the fault in the systems provide input to a fault-tolerant control system that ensures optimal operation of the gas lift system.

Conflict of interest

There is no conflict of interest in this article.

References

[1] Aamo, O.M., Eikrem, G.O., Siahaan, H.B. and Foss, B.A., 2005. Observer design for multiphase flow in vertical pipes with gas-lift—theory and experiments. *Journal of process control*, *15*(3), pp.247-257.

[2] Diehl, F.C., Almeida, C.S., Anzai, T.K., Gerevini, G., Neto, S.S., Von Meien, O.F., Campos, M.C., Farenzena, M. and Trierweiler, J.O., 2018. Oil production increase in unstable gas lift systems through nonlinear model predictive control. *Journal of Process Control*, *69*, pp.58-69.

[3] Eikrem, G.O., Aamo, O.M. and Foss, B.A., 2008. On instability in gas lift wells and schemes for stabilization by automatic control. *SPE Production & Operations*, *23*(02), pp.268-279.

[4] Eikrem, G.O., Foss, B., Imsland, L., Hu, B. and Golan, M., 2002. Stabilization of gas lifted wells. *IFAC Proceedings*, *35*(1), pp.139-144.

[5] Eikrem, G.O., Imsland, L. and Foss, B., 2004. Stabilization of gas lifted wells based on state estimation. IFAC Proceedings Volumes, 37(1), pp.323-328.

[6] Garcia, A.P., 2013, November. Stability analysis and stabilization of gas lift systems. In *22nd International Congress of Mechanical Engineering*, November (pp. 3-7).

[7] Jahanshahi, E., 2013. Control Solutions for Multiphase Flow: Linear and nonlinear approaches to anti-slug control. Norges teknisk-naturvitenskapelige universitet, Fakultet for naturvitenskap og teknologi, Institutt for kjemisk prosessteknologi.

[8] Scibilia, F., Hovd, M. and Bitmead, R.R., 2008. Stabilization of gas-lift oil wells using topside measurements. *IFAC Proceedings 41*(2), pp.13907-13912.

[9] Shi, J., Al-Durra, A., Errouissi, R. and Boiko, I., 2019. Stabilization of artificial gas-lift process using nonlinear predictive generalized minimum variance control. Journal of the Franklin Institute, 356(4), pp.2031-2059.

[10] Adukwu, O. and Oku, D.E., 2018. Modeling, Analysis and Monitoring DC Motor Using Proportional-Integral Controller and Kalman Filter. *Methodology*, *1*, p.2.

[11] Adukwu, O., Odloak, D., Saad, A.M. and Junior, F.K., 2022. State Estimation of Gas-Lifted Oil Well Using Nonlinear Filters. *Sensors*, *22*(13), p.4875.

[12] Basseville, M. and Nikiforov, I.V., 1993. Detection of abrupt changes: theory and application (Vol. 104). Englewood Cliffs: Prentice Hall.

SciEn
Journal of Engineering Advancements Vol. 04(02) 2023, pp 40-48 https://doi.org/10.38032/jea.2023.02.002

Determination of Production System Effectiveness Based on Sustainable Global Standards

A. J. Yakubu[1,*], *B. Kareem*[2], *and B. O. Akinnuli*[2]

[1]Department of Civil Engineering, Federal Polytechnic Ile-Oluji, Nigeria
[2]Department of Industrial and Production Engineering, Federal University of Technology, Akure, Nigeria.

Received: May 17, 2023, Revised: June 18, 2023, Accepted: June 18, 2023, Available Online: June 27, 2023

ABSTRACT

Production system effectiveness determine to measure the sustainability of the established industries demands the development of a model for resolving global sustainable productivity challenges. The attributes (internal and external) of industrial failure were determined using questionnaire administration and oral interviews of industry experts in five (5) selected production companies in Nigeria: (Company A); (Company B); (Company C); (Company D) and (Company E). Production System Effectiveness (PSE) factors: Availability A, Performance P, and Quality Q were determined to arrive at manageable decision-making criteria under uncertainty, risk, or competition. Initial measures of PSE were based on the input internal factors (manpower, machine, material, energy, management, information/communication, money, and marketing), while sustainability decisions were determined using globally acceptable standards. The model was tested using data (weighted and normal) from the stated companies to determine their sustainability performances, while paired t-test statistic was used to test the levels of significant difference between weighted (WPSE) and normal (PSE) at 5%. The results indicated varying optimum decisions which were influenced by the nature/types of competition, risk, and standard of measure. The statistical result showed that there was a significant difference between the PSE and WPSE. These differences had little or no effect on optimum decision-making in all companies investigated.

Keywords: Global, Standard, Production, Effectiveness, Sustainability.

1 Introduction

Sustainability means meeting the needs without compromising the production system challenges [1]. Apart from material resources, machinery, manpower, energy, marketing, information technology, and money/funding sustainability are very important. Sustainable productivity performance of industries required optimum harmonization of the stated resources in the delivery of the core production process [2].

Efforts and programs targeted at improving productivity in Nigerian industries have not yielded significant results [3]. The distribution of labor productivity by age of firms reveals that an average productivity of NGN10, 198 per worker is recorded among firms above 20 years of age and that labor productivity increases with the firm's age [4].

With increasing globalization, human capital and manpower development, machine revolution, material advancement, modern communication, advanced marketing, and energy hybridization, a good sources utilization policy is required and can be accessed through qualitative education and training in sources management [5].

Human capital development is crucial and ultimate in propelling productivity. Technological advancements are products of human minds and can only be made productive by positive monitoring and competitiveness. The energy sector also contributed to the industrialization of many nations; its failure however has an advert effect on the nation's economic growth [4].

The manufacturing industry constitutes a large impact that undertakes a series of activities, which include the production of different items, machines, equipment, etc. There are a range of sections in the manufacturing industry, from the managerial section down to production, maintenance, and inspection departments. Due to the competition between corporations, industries, businesses, firms, and organizations, there is always the desiring need for something new. Every industry or firm must have competent management in place to ensure that the production process is always on the right track. For the manufacturing industries to compete favorably with one another, they must be innovative [6]. A competitive manufacturing industry is an ingredient of sustainable development [7]. Sustainable development also means attaining a balance between environmental protection and human capacity building and between present and future needs [8]. In all cases, manufacturing (production) industries played a very important role in achieving a sustainable development goal by 2030 [3].

Production industries required a good transportation system (by land, water, or air) which comprised automobiles, marines, and aeronautics. Transportation industries have played a good role in sustainable development in the areas of safe transportation of raw materials and finished goods to/from the production industries [9]. A good transportation system has enabled waste elimination, and prompt availability of raw materials and other production resources as and when required for production activities, thereby improving resource utilization, procurement management, and sustainability [10]-[12].

Developing countries need accelerated growth and the manufacturing industry provides the bulk of this transition to developed economies. This means a bulk of investment is

*Corresponding Author Email Address: yakubqiblah@yahoo.com

necessary to develop infrastructure for the industries to thrive, reach their sustainable capacities, and attain accelerated Gross Domestic Product (GDP). On this basis, strategic planning geared towards promoting adequate investment in the manufacturing industry is necessary [13].

The global demand for effective utilization of both humans and machinery is increasing due to wastage incurred during product manufacturing. Excessive waste generation has made entrepreneurs find it difficult to break even. The development of a dynamic error-proof Overall Equipment Effectiveness (OEE) model for optimizing the operations of a complex production system is targeted at minimizing/eradicating generated wastes/losses [3].

There is a high need to move from ordinary mechanization to automation of industrial processes to improve efficiency [14]. Maximum productivity is highly needed hence, automation must be followed by a lean workforce [15].

The global mantra in the past four decades has culminated in the desire to achieve sustainability and sustainable development. This mantra has stemmed from concerns for the future, in terms of resource endowment, human health, and the environment. Nigeria has yet to meet this goal as there are several challenges to sustainable industrial development [16]. The progressive incorporation of information and communication technologies (ICT) [17] and their combination with production process technologies have made manufacturing operations more intelligent and sustainable [18].

Interestingly, the research that focused on contributing to sustainability was mainly dedicated to Europe, the Far East, and the Southeast regions of Asia. One reason for this result was huge EU funding for research, which meant that Europe's researchers had access to huge funds for industrial sustainability and digitalization. This funding initiative had contributed to the digital transformation of industries in Europe [18].

In advanced technology platforms, data are a crucial factor in promoting sustainable production and supply chain operations [19]. Sustainable manufacturing is positively mediated using sensors, intelligent algorithms, and actuators to permit data collection in the manufacturing environment [20]. Historic product characteristics can be saved in the blockchain, which allows users to identify the origin, quality, and lead time of the product [21]-[22]. Smart supply chains and transportation systems are critical to industrial productivity [23]-[24].

Design change propagation is a primary source of risk and innovation in complex product (CP) development of a production system which can affect processing sustainability [25]. Linking product design to customer behavior is a good factor in sustaining productivity in a dynamic production environment. The simulation concept was also used in making decisions as it affect productive processes in the past. However, with the availability of solutions and technologies, simulation is no longer a tool with limited scope and analysis. Artificial intelligence with physical systems was considered to allow virtual models to be sensitive to physical changes and aligned with the current state of production processes [26].

The application of an innovative energy system is capable of resolving the challenges created by the inadequacy of energy in the production system [27]-[28]. The effectiveness of operational level and management (EM) practices and their long-term impacts on material inventory was assessed using data from U.S. industrial facilities [29]. Demand-side mitigation solutions such as changing peoples' consumption behaviors can substantially help limit climate change. In the manufacturing realm, promoting, and directing the consumption behavior of customers is a good factor in encouraging sustainable industrial development [30]. Sustainability measures were being re-designed to provide a measurement of the production system within the link of accountability [1]. Measuring and evaluating the performance of production process sustainability is still not a common practice in some companies [31].

1.1 Production System Effectiveness (PSE)

PSE depends on availability rate, performance efficiency, and quality rate. Therefore, PSE increases with the increase of these three elements. An increase in availability rate reduces buffer inventories needed to protect downstream production from breakdowns and increases effective capacity. An increase in the rate of quality products means that there is less scrap and rework, reduces costs, and yields a higher rate of quality [32]-[33]. PSE is a complete performance measurement indicator, but to make it real it requires modification in terms of weights allocation [34], inclusion of production system dynamism, and consideration of production competitiveness. Factors affecting PSE are not equally important in all aspects and different weights of elements played a critical role.

Wudhikarn *et al.* [35] proposed a new PSE indicator without considering production competitiveness. PSE and weighted PSE measures are more realistic because of production dynamism, corruption, and competitiveness consideration.

Global sustainable standards in which production system effectiveness/productivity were been measured are enumerated in Table 1. In this study the choice of sustainable PSE was based on standards, this was rare in the past studies.

Table 1 Sustainable standard of production system effectiveness/productivity

Sustainable Standards/Classes	Effectiveness/Productivity Range	Sustainability Implication
Global Standard, P(G)	≥ 0.85	Sustainable , [36]-[37]

2 Methodology

2.1 Model Development

Factors that hindered productivity in terms of availability, quality, and performance in selected production systems were identified using questionnaire administration and oral interviews with manufacturing experts. These productivity challenges were caused by both external (outside production system) and internal (within production system) factors, individually and collectively. The identified internal factors are manpower, money, machine, energy, management, information/communication, material, and marketing while external factors are sustainable development trends, industrial revolution, and globally sustainable/acceptable standards. The block diagram that shows the relations among the internal, external, and production system effectively is given in Fig. 1.

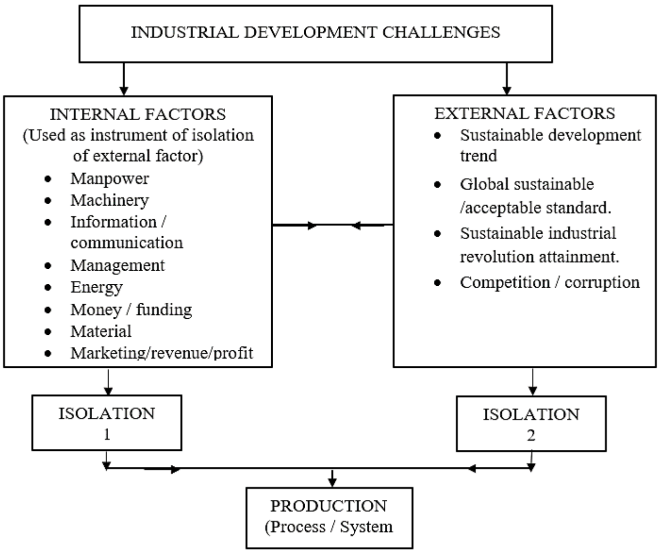

Fig. 1 Block Diagram of Production (Process / System)

The challenges posed by the internal/external factors hindered the attainment of the maximum obtainable productivity index of unity (1). That is for the N number of internal factors, productivity continued to decrease with an increased number of factors N called challenges. Therefore, traditionally, productivity or production system effectiveness (PSE) was mathematically presented as Eq. (1) was modified as in Eq. (2) to take care of the stochastic nature of the process.

$$PSE = APQ \tag{1}$$

$$PSE = P(S) = APQ \tag{2}$$

where,

PSE is production system effectiveness,
A is Availability, P is Performance, and Q is Quality.

Eqs. (1) and (2) are similar because their outcome is always less than 1 but they are different because the former is static while the latter is probabilistic, its outcome can change in space and time. This indicated the real nature of the production system. On this basis, Eq. (2), on consideration of the stated challenges was modified as Eq. (3).

$$PSE = P(S) = APQ < 1 \tag{3}$$

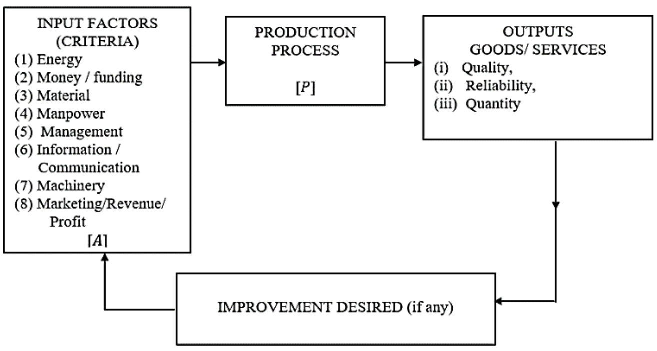

Fig. 2 Block Diagram of proposed Model Characteristic

The problem at hand is how to improve productivity such that external factors hindrance is mitigated. That is a globally accepted standard, $P(G)$, which is termed exogenous variables are satisfied as presented in Eq. (4).

$$P(S) = APQ \geq P(G) \tag{4}$$

Where, P(G) is global acceptable standard.

The block diagram shown in Fig. 2 depicted the improvement strategy developed at meeting the set standards, with the main objective of meeting the condition of productivity stated in Eq. (5).

$$P(S) = APQ = 1 \tag{5}$$

Eq. (5) was made robust to allow weighting of the production system effectiveness factors using the Rank-Order Centroid (ROC) method [34] in which rank 1, 2, and 3 were allocated for lowest, average and highest weights respectively to enable its application in all categories of production system productivity measures. The weighting production system effectiveness (WPSE) was calculated as stated in Eq. (6). Eq. (7) gives the general equations for assigning weight ranking which results to w_1, w_2 and w_3 as stated in Eqs. (8), (9), and (10) respectively [34].

$$P(S) = WPSE = w_1 I + w_2 P + w_3 O = 1 \tag{6}$$

Based on the ROC method [34]:

$$W_i = \left(\frac{1}{K}\right) \sum_{j-i}^{K} \frac{1}{r_k} \tag{7}$$

$$w_1 = \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots \ldots . +1/K\right)/K \tag{8}$$

$$w_2 = \left(0 + \frac{1}{2} + \frac{1}{3} + \cdots \ldots . +1/K\right)/K \tag{9}$$

$$w_3 = (0 + 0 + 0 + \cdots \ldots . +1/K)/K \tag{10}$$

where:

r_k is the rank of the k^{th} objective
K is the total number of objectives
w_i is the normalized approximate ratio scale weight of the i^{th} objective.
w_1 is weight of availability P(I) attribute
w_2 is weight of performance P(p) attribute
w_3 is weight of quality P(O) attribute

It is inferable from the stated Eqns. that if,

$$P(S) = 1, \text{ no challenges in the system (fully sustainable)} \tag{11}$$

$$P(S) = 0, \text{ System has collapsed} \tag{12}$$

$$P(S) < 0{,}85, \text{ System is gradually collapsing but may be sustainable} \tag{13}$$

These outcomes (Eqs. (11)-(13)) led to the establishment of two major decision variables (sustainable or unsustainable, under three conditions (fully sustainable/unsustainable, fairly/averagely sustainable, and fully unsustainable/sustainable), respectively. These alternative decision outcomes are shown in Fig. 3.

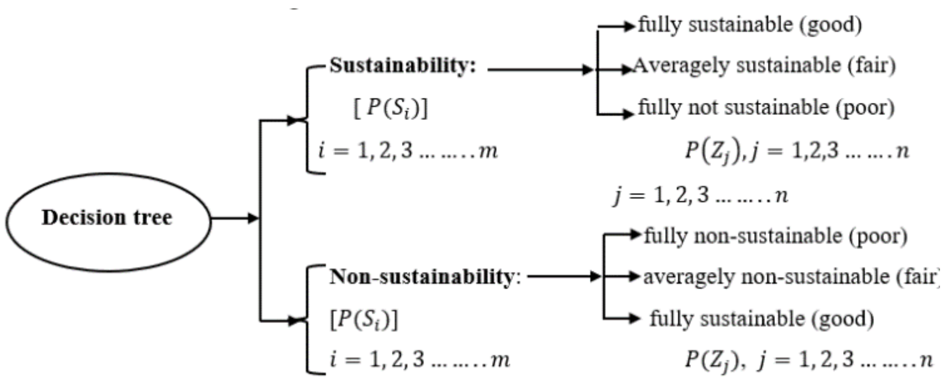

Fig. 3 Decision Tree on production process sustainability

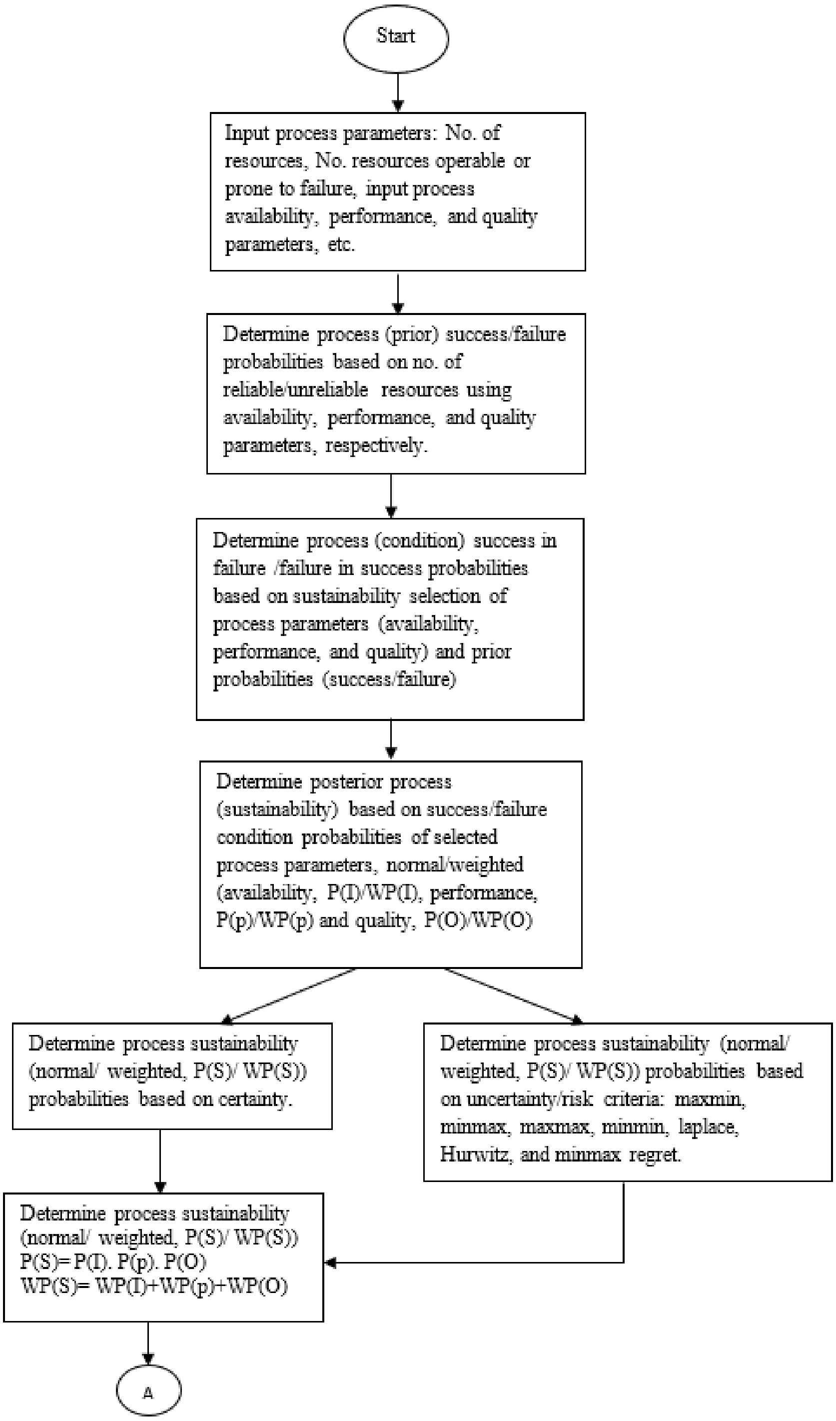

Fig. 4 System development flowchart

First, Availability A, performance P, and Quality Q productivity measures were modified to reflect the real and dynamic probabilistic situation of the production system (as probabilities of input resources availability A, process performance P and output quality Q. Second, outcomes from the first step were partitioned into either success (good), $P(\,)_S^*$, or failure (poor), $P(\,)_f^*$ productivity. Third, the binomial probability model was modified and applied to translate the process into three real-life productivity scenarios: good or sustainable; fairly or averagely sustainable; and poor or unsustainable. Forth, prior probabilities of process sustainability were measured based on the functionality of available production resources by focusing on radical production machinery. Fifth, process conditional probability was estimated based on success, failure, and success/failure sustainability scenarios. Sixth, process sustainability (posterior) probability, $P[S_i/Z_j]$ was established under normal and weighting for availability, performance, and quality, S_i respectively at a given condition, Z_j good, poor, or both. Next, Production System Effectiveness $PSE/P(S)$ was determined under normal and weighting conditions.

2.2 Flowchart for Computer Software programming

Flowchart (Fig. 4) of the Production Process starts from the identification of Production System Effectiveness (PSE) factors: Availability$A$, Performance P and Quality Q. The PSE using Traditional approach (APQ) for measuring the production system effectiveness based on the production input (internal) factors (manpower, machine, material, energy, management, information/communication, money, and marketing).

2.3 Model Validation and Performance Test

Data were collected to test the efficacy of the model. The model was tested using the first round of data collected (70 % of the data) while the second round of the data (30 %) was used for validation. Paired T-test statistics were used to test if there existed a significant difference between PSE (μ_2) and WPSE (μ_2) for a given production system (company).

Hypothesis:

$H_O: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

Decision rule: reject H_0 if $P_{calculated}$<p-value.

Inference: Since P_{cal}<p-value. There is enough evidence to reject.

For industry $0.1 \geq 1$, it is expected that time losses due to failure / idle time should not exceed the minimum range of Availability, Performance, and Quality as established, in past studies [3], [38].

i. 0.1 – 0.50% for Industry 1.0 (poor, not sustainable)
ii. 0.51 – 0.84% for Industry 2.0 to 3.0 (fair, averagely sustainable)
iii. 0.85 – 1.0% for Industry 4.0 (excellent, and sustainable)
iv. Greater than or equal to 1.0 for Industry 5.0 (outstanding, super sustainable)

2.4 Collection of Data

Relevant data were collected using a questionnaire and oral interview conducted in five (5) selected companies, labeled A, B, C, D, and E. The data were collected on the production process, working hours, downtime, product rejection, etc. These data were used for estimating relevant parameters as contained in the developed model. Estimated parameters include Availability Rate, Production Process Performance, Quality rate, Overall Production System Effectiveness (PSE), and decisions on production system process sustainability were made based on $P(G)$, criteria established from the literature. The summaries and nature of the data collected from Company A were given in Table 3 and the same method was used for collection of data for Company B, C, D, and E. In addition, data collection on weights ranking on Production Process Effectiveness factors were also given in Table 4.

Table 2 Summary of the Mathematical Model Development

S/n	Parameter	Traditional / Convectional (Old Model)	Definition of symbols
1	Initial condition of the production process	Availability A= $\frac{\text{Operation time}}{\text{Loading time}}$	A=Availability
		Performance P= $\frac{\text{Net Processing time}}{\text{Operating time}}$	P=Performance
		Quality Q = $\frac{\text{Processed amount} - \text{defect amount}}{\text{Processed amount}}$	Q=processed amount
2	Sustainability evaluation		P(G),Global Acceptable standard ≥0.85, 1.0

Table 3 Data Collection on Availability, Performance, and Quality of Company A

Company A											
Process line Product: Cement processing line/Eight (8) hours shift											
Input Factor	Availability P(I) /hour = A		Performance P(p) /hour = P				Quality P(O)/quantities (kg) = Q				
	Plants Time/ (Set-up / h)	Loading Time = (Process + loading + off-loading) time /h	Process Time /h	Operating time (Cycle time)/h			Processed Amount/kg	Defect loses amount/kg			
				Idling losses/h	Minor stoppage /h	Reduced speed /h		Rework losses	Defect losses	Start-up loses	Scrapped loses
Manpower	8	8	8	1	2	0.5	1,200	25	10	5	2
Machinery	6	8	7	1	2	1	1,000	50	22	12	3
Info. /Comm	8	8	8	0.5	1	1	950	15	5	5	1
Management	6	8	7	0.5	0	1	700	20	14	5	2
Energy	7	8	6	0.5	0	3	1500	22	12	5	4
Money/fund	8	8	7	0.5	0	1	2000	50	15	7	5
Material	8	8	7	1	0.5	0.5	1150	12	20	20	7
Marketing	8	8	8	0.5	1	0.5	1100	11	20	18	2
$\boldsymbol{PSE = APQ}$	**0.9210**			**0.8806**			**0.9890**				
	$(0.9210 \times 0.8806 \times 0.9890) = \mathbf{0.8021}$										

Table 4 Data Collection on Weights Ranking on Production Process Effectiveness Factors

Attributes PSE	Ranking $(\boldsymbol{r_k})$	Numerical calculation	Weight
Company A			
P(I)	1	$W_1 = (1 + 1/2 + 1/3)/3$	0.61
P(p)	3	$W_2 = (1/3)/3$	0.11
P(O)	2	$W_3 = (1/2 + 1/3)/3$	0.28
Company B			
P(I)	3	$W_2 = (1/3)/3$	0.11
P(p)	1	$W_2 = (1 + 1/2 + 1/3)/3$	0.61
P(O)	2	$W_3 = (1/2 + 1/3)/3$	0.28
Company C			
P(I)	2	$W_1 = (1/2 + 1/3)/3$	0.28
P(p)	3	$W_2 = (1/3)/3$	0.11
P(O)	1	$W_3 = (1 + 1/2 + 1/3)/3$	0.61
Company D			
P(I)	1	$W_1 = (1 + 1/2 + 1/3)/3$	0.61
P(p)	3	$W_2 = (1/3)/3$	0.11
P(O)	2	$W_3 = (1/2 + 1/3)/3$	0.28
Company E			
P(I)	2	$W_1 = (1/2 + 1/3)/3$	0.28
P(p)	1	$W_2 = (1 + 1/2 + 1/3)/3$	0.61
P(O)	3	$W_3 = (1/3)/3$	0.11

3 Results and Discussion

A summary of the normal and weighted Production System Effectiveness (PSE and WPSE) results under traditional (APQ) was presented in Table 5. It can be revealed that the traditional approach under equal weights has not produced sustainable outcomes in all companies investigated, while companies A, D, and E had sustainable performance under the weighted arrangement. In this case, production system effectiveness was sustainable in all companies in both normal and weighted scenarios.

Table 5 and Fig. 5, show the summary of the Production System Effectiveness, PSE; Weighted Production System Effectiveness, WPSE; and the corresponding decision outcomes (sustainable, fairly sustainable, or unsustainable) for companies A, B, C, D, and E, under competitive production environment with reference to global acceptable standard factor.

Table 6 shows that only the Maximax criterion was sustainable (D_S) on Production System Effectiveness (PSE) and Weighted Production System Effectiveness (WPSE) which assumed no presence of competition. Also, Laplace and Hurwitz's criteria were sustainable (D_F) on WPSE only with the presence of fair competition. Maximin, Minimax, Minimin, and Minimax Regret criteria were unsustainable (D_U) on PSE and WPSE which assumed that full competition was in place. Therefore, company A can only survive under the Maximax criterion that is without competition. Hypothesis test (paired T-test) results $p_{cal} = 0.007$, p-value 0.05 ($p_{cal} < p_{-value}$) between PSE and WPSE indicated that there was a significant difference between the normal Production System Effectiveness (PSE) and weighted Production Effectiveness (WPSE) at 5% level of significance.

Similar decision outcomes were obtained for company B with little improvement. There were better decision outcomes in terms of sustainable productivity in Company C as a majority of the good decisions fell under either sustainable or sustainable processes. However, PSE and WPSE results were significantly different at the 5 % level.

Decision results from Company D indicated that the company cannot sustain productivity under keen competition. The decision results from Company E were very close to that of Company D, with similar significant difference characteristics between PSE and WPSE. In all cases, however, there was no wide gap in overall decision-making related to the PSE and WPSE outcomes.

4 Conclusion

A model capable of resolving sustainable productivity challenges of production industries was established in this study. The model was tested using data obtained from five Nigerian Companies (Company A, B, C, D, and E). Production System Effectiveness (PSE) factors: Availability P(I), Performance P(p), and Quality P(O) were determined using the Modified Bayesian Approach (MBA) to arrive at manageable decision-making criteria under a normal and/or competitive production environment. The results obtained from the model revealed that varying system sustainability decision-making was due to competitiveness and standard of measure. There was a statistically significant difference between the PSE and WPSE in many industrial cases tested, but these differences had little or no effect on optimum decision-making in all companies investigated.

Decision outcomes were obtained for companies A and B with little improvement. There were better decision outcomes in terms of sustainable productivity in Company C as the majority of the good decisions fell under either sustainable or sustainable processes. However, PSE and WPSE results were significantly different at the 5 % level.

Company D indicated that the company cannot sustain productivity under keen competition. The decision results from Company E were very close to that of Company D, with similar significant difference characteristics between PSE and WPSE. In all cases, however, there was no wide gap in overall decision-making related to the PSE and WPSE outcomes.

Table 5 Normal Production System Effectiveness (PSE)

Company	Conventional/ Traditional Approach (APQ) (normal PSE, and weighed WPSE	Industrial revolution standards	Sustainability Global limit	Sustainability Revolution class
	$PSE = APQ$	≥0.85, 1.0		
A	0.8016	I2.0	0.51 – 0.84%	Fairly sustainable
B	0.4849	I2.0	0.1 – 0.50%	Not sustainable
C	0.3430	I1.0	0.1 – 0.50%	Not sustainable
D	0.6970	I2.0	0.51 – 0.84%	Fairly sustainable
E	0.7128	I2.0	0.51 – 0.84%	Fairly sustainable

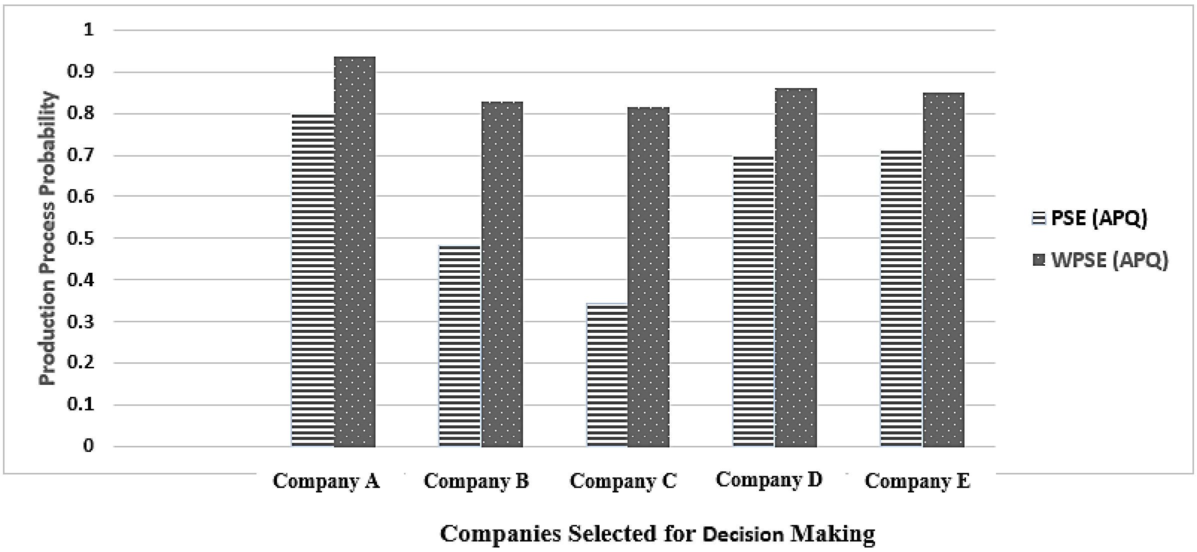

Fig. 5 PSE and WPSE Comparison under Conventional Approach

Table 6 PSE and WPSE Sustainable Decision Analysis under Competitive (Uncertainty)

Company Type	Decision-Making Criteria													
	Maximin		Minimax		Maximax		Minimin		Laplace		Hurwicz		Minimax Regret	
	PSE	$WPSE$	PSE	$WPSE$	PSE	$WPSE$	PSE	$WPSE$	PSE	$WPSE$	PSE	$WPSE$	PSE	$WPSE$
Company A	0.0176	0.4721	0.1240	0.4088	0.9230	0.9583	0.0001	0.0256	0.4255	0.8102	0.3219	0.7778	0.0904	0.3617
Company B	0.0008	0.1873	0.3978	0.8127	0.9761	0.9947	0.0000	0.0034	0.3470	0.6857	0.2324	0.5910	0.3820	0.8074
Company C	0.0004	0.0768	0.7863	0.9230	0.9949	0.9983	0.0000	0.0017	0.2744	0.6501	0.1551	0.5376	0.7820	0.9213
Company D	0.0000	0.1512	0.3445	0.8488	0.9716	0.9949	0.0000	0.0033	0.1910	0.4958	0.2726	0.5731	0.3276	0.8438
Company E	0.0000	0.0222	0.8758	0.9778	0.9963	0.9995	0.0000	0.0007	0.1402	0.5505	0.1431	0.5108	0.8507	0.9771
Decision:														
$P(T), \geq 0.85$	D_U	D_U	D_S (E)	D_S (C, D, E)	D_S	D_S	D_U	D_U	D_U	D_U	D_U	D_U	D_S (E)	D_S (C, D)
$P(G) \geq 0.85$	D_U	D_U	D_S (E)	D_S (C, D, E)	D_S	D_S	D_U	D_U	D_U	D_U	D_U	D_U	D_S (E)	D_S (C, D)
$P(R) = 0.1 - 0.5$, $(I1.0 - I2.0)$	D_U	D_F (A, B, D)	D_F (A, B, D)	D_S (A)	D_S	D_S	D_U	D_U	D_F	D_F (D)	D_S	D_S	D_F (B, C)	D_F (A)
$P(R) = 0.51 - 0.84$, $(I2.0 - I3.0)$	D_U	D_U	D_S (C)	D_S (B)	D_S	D_S	D_U	D_U	D_U	D_F (A, B, C, E)	D_U	D_S	D_F (C)	D_S (B, D)
$P(R) = 0.85 - 1.0$, $(I4.0 - I5.0)$	D_U	D_U	D_S (E)	D_S (C, D, E)	D_S	D_S	D_U	D_U	D_U	D_U	D_U	D_U	D_S (E)	D_S (C, E)

Decision:
D_S is (Sustainable)
D_F is (Fairly sustainable)
D_U is (unsustainable)
P(T) is sustainable trend
P(G) is global acceptable standard
P(R) is industrial revolution
I is industrial revolution

Acknowledgement

The authors acknowledge the Tertiary Education Trust Fund (TETFund) in Nigeria for sponsoring this research work.

References

[1] Atkins, J., Doni, F., Gasperini, A., Artuso, S., La Torre, I. and Sorrentino, L., 2022. Exploring the effectiveness of sustainability measurement: which ESG metrics will survive COVID-19?. Journal of Business Ethics, pp.1-18

[2] Wudhikarn, R. and Manopiniwes, W., 2010, June. Autonomous maintenance using total productive maintenance approach: A case study of synthetic wood plank factory. In Proc. Technology Innovation & Industrial Management Conf.

[3] Kareem, B., Alabi, A.S., Ogedengbe, T.I., Akinnuli, B.O. and Aderoba, O.A., 2020. Development of OEE Error-Proof (OEE-EP) Model for Production Process Improvement. The Journal of Engineering Research [TJER], 17(2), pp.59-74.

[4] Agbu, O., 2007. The iron and steel industry and Nigeria's Industrialization: Exploring cooperation with Japan. (No Title). Agbu, O. (2007).

[5] World Bank, 2006. Investment Climate Survey Data. Washington, DC: World Bank, 1(1), pp.1-21.

[6] Fayomi, O.S.I., Akande, I.G., Loto, R.T. and Ongbali, S.O., 2019, July. Sustainable need in manufacturing industry in Nigeria toward quality, policy and planning. In AIP Conference Proceedings (Vol. 2123, No. 1, p. 020072). AIP Publishing LLC.

[7] Dal, B., Tugwell, P. and Greatbanks, R., 2000. Overall equipment effectiveness as a measure of operational improvement–a practical analysis. International Journal of Operations & Production Management, 20(12), pp.1488-1502.

[8] Singh, V., 2014. An impact and challenges of sustainable development in global era. Journal of Economics and Development Studies, 2(2), pp.327-337.

[9] Kareem, B. and Jimoh, Y.A., 2017. Modelling Failure Rate of Automobile Crankshafts based on Distance Travelled and Age.

[10] Shahbazi, S., Salloum, M., Kurdve, M. and Wiktorsson, M., 2017. Material efficiency measurement: empirical investigation of manufacturing industry. Procedia Manufacturing, 8, pp.112-120.

[11] Silaipillayarputhur, K., 2018. Process safety management in manufacturing industries a review. International Journal of Engineering & Technology, 7(2), pp.540-543.

[12] Jain, V.N. and Mishra, D., 2018. Blockchain for supply chain and manufacturing industries and future it holds. International Journal of Engineering Research, 7(9), pp.32-39.

[13] Mićić, V. and Janković, N., 2017. Investment in the manufacturing industry of Serbia. Bankarstvo, 46(4), pp.52-73.

[14] Singh, B.N., 2018. Role of Automation in Steel Industry. no. May.

[15] Landry, J. and Ahmed, S.A., 2016. Adoption of leanness in the manufacturing industry. Univers. J. Manag, 4(1), pp.1-4.

[16] Momodu, A.S., Aransiola, E.F., Okunade, I.D., Ogunlusi, G.O., Awokoya, K.N., Ogundari, I.O., Falope, O.T., Makinde, O.W. and Akinbami, J.F.K., 2019, May. Greening Nigeria's economy for industrial and environmental sustainability: Polyurethane production as a test case. In Natural Resources Forum (Vol. 43, No. 2, pp. 73-81). Oxford, UK: Blackwell Publishing Ltd.

[17] Felsberger, A., Qaiser, F.H., Choudhary, A. and Reiner, G., 2022. The impact of Industry 4.0 on the reconciliation of dynamic capabilities: Evidence from the European manufacturing industries. Production Planning & Control, 33(2-3), pp.277-300.

[18] European Commission Horizon [ECH], 2020. Work Programme 2018–2020. Available online:https://ec.europa.eu/commission/presscorner/detail/en/ MEMO, 17(8), pp.1-14.

[19] Culot, G., Orzes, G., Sartor, M. and Nassimbeni, G., 2020. The future of manufacturing: A Delphi-based scenario analysis on Industry 4.0. Technological forecasting and social change, 157, p.120092.

[20] Garrido-Hidalgo, C., Hortelano, D., Roda-Sanchez, L., Olivares, T., Ruiz, M.C. and Lopez, V., 2018. IoT heterogeneous mesh network deployment for human-in-the-loop challenges towards a social and sustainable Industry 4.0. Ieee Access, 6, pp.28417-28437.

[21] Pournader, M., Shi, Y., Seuring, S. and Koh, S.L., 2020. Blockchain applications in supply chains, transport and logistics: a systematic review of the literature. International Journal of Production Research, 58(7), pp.2063-2081.

[22] Gan, T.S., Steffan, M., Grunow, M. and Akkerman, R., 2022. Concurrent design of product and supply chain architectures for modularity and flexibility: process, methods, and application. International Journal of Production Research, 60(7), pp.2292-2311.

[23] Zheng, F., Wang, Z., Zhang, E. and Liu, M., 2022. K-adaptability in robust container vessel sequencing problem with week-dependent demands of a service route. International Journal of Production Research, 60(9), pp.2787-2801.

[24] Shajalal, M., Hajek, P. and Abedin, M.Z., 2023. Product backorder prediction using deep neural network on imbalanced data. International Journal of Production Research, 61(1), pp.302-319.

[25] Li, R., Yi, H. and Cao, H., 2022. Towards understanding dynamic design change propagation in complex product development via complex network approach. International Journal of Production Research, 60(9), pp.2733-2752.

[26] dos Santos, C.H., Montevechi, J.A.B., de Queiroz, J.A., de Carvalho Miranda, R. and Leal, F., 2022. Decision support in productive processes through DES and ABS in the Digital Twin era: a systematic literature review. International Journal of Production Research, 60(8), pp.2662-2681.

[27] Bevan, D., Collier, P. and Gunning, J.W., 1999. The political economy of poverty, equity and growth: Nigeria and Indonesia. Oxford University Press.

[28] McKone, K.E., Schroeder, R.G. and Cua, K.O., 2001. The impact of total productive maintenance practices on manufacturing performance. Journal of operations management, 19(1), pp.39-58.

[29] Zu, X. and Cong, Y., 2022. Green at source: an empirical examination of the effectiveness and sustainability of operational-level environmental management practices in US industry. Total Quality Management & Business Excellence, 33(11-12), pp.1213-1232.

[30] Majer, J.M., Henscher, H.A., Reuber, P., Fischer-Kreer, D. and Fischer, D., 2022. The effects of visual sustainability labels on consumer perception and behavior: A systematic review of the empirical literature. Sustainable Production and Consumption.

[31] Baumer-Cardoso, M.I., Ashton, W.S. and Campos, L.M., 2023. Measuring the adoption of circular economy in manufacturing companies: the proposal of the Overall Circularity Effectiveness (OCE) index. Circular Economy and Sustainability, 3(1), pp.511-534.

[32] Jonsson, P. and Lesshammar, M., 1999. Evaluation and improvement of manufacturing performance measurement systems-the role of OEE. International Journal of Operations & Production Management.

[33] Fredendall, L.D., Patterson, J.W., Kennedy, W.J. and Griffin, T., 1997. Maintenance: modeling its strategic impact. Journal of managerial issues, pp.440-453.

[34] Raouf, A., 1994. Improving capital productivity through maintenance. International Journal of Operations & Production Management.

[35] Wudhikarn, R., Smithikul, C. and Manopiniwes, W., 2010. Developing overall equipment cost loss indicator. In Proceedings of the 6th CIRP-Sponsored International Conference on Digital Enterprise Technology (pp. 557-567). Springer Berlin Heidelberg.

[36] Almeanazel, O.T.R., 2010. Total productive maintenance review and overall equipment effectiveness measurement. Jordan Journal of Mechanical and Industrial Engineering, 4(4).

[37] Binti Aminuddin, N.A., Garza-Reyes, J.A., Kumar, V., Antony, J. and Rocha-Lona, L., 2016. An analysis of managerial factors affecting the implementation and use of overall equipment effectiveness. International journal of production research, 54(15), pp.4430-4447.

[38] Mohammed, S.A., 2008, March. Privatization of the iron and steel industry in Africa. In 8th International Arab Iron and Steel Conference, held at Doha, Qatar 17th–19th march.

SciEn
Journal of Engineering Advancements Vol. 04(02) 2023, pp 49-56 https://doi.org/10.38032/jea.2023.02.003

Detecting Level of Depression from Social Media Posts for the Low-resource Bengali Language

Md. Nesarul Hoque[1,*] *and Umme Salma*[2]

[1] Department of Computer Science and Engineering, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj-8100, Bangladesh,
[2] Department of Computer Science & Engineering, Bangladesh University, Dhaka-1207, Bangladesh

Received: May 17, 2023, Revised: June 19, 2023, Accepted: June 19, 2023, Available Online: June 28, 2023

ABSTRACT

Depression is a mental illness that suffers people in their thoughts and daily activities. In extreme cases, sometimes it leads to self-destruction or commit to suicide. Besides an individual, depression harms the victim's family, society, and working environment. Therefore, before physiological treatment, it is essential to identify depressed people first. As various social media platforms like Facebook overwhelm our everyday life, depressed people share their personal feelings and opinions through these platforms by sending posts or comments. We have detected many research work that experiment on those text messages in English and other highly-resourced languages. Limited works we have identified in low-resource languages like Bengali. In addition, most of these works deal with a binary classification problem. We classify the Bengali depression text into four classes: non-depressive, mild, moderate, and severe in this investigation. At first, we developed a depression dataset of 2,598 entries. Then, we apply pre-processing tasks, feature selection techniques, and three types of machine learning (ML) models: classical ML, deep-learning (DL), and transformer-based pre-trained models. The XLM-RoBERTa-based pre-trained model outperforms with 61.11% F1-score and 60.89% accuracy the existing works for the four levels of the depression-class classification problem. Our proposed machine learning-based automatic detection system can recognize the various stages of depression, from low to high. It may assist the psychologist or others in providing level-wise counseling to depressed people to return to their ordinary life.

Keywords: Depression, Machine learning, Multi-class Classification, Low-resource Language, Bengali.

1 Introduction

The mood, thoughts, behavior, and physical health of a person all affect the mental health illness known as depression. It can affect a person's capacity to function in their way of life and range in intensity from minor to severe. Aside from causing emotional pain, it can also have negative effects on one's physical health. It frequently carries a greater mortality rate as well as a higher chance of developing chronic diseases like obesity, diabetes, and cardiovascular disease. Depression can arise from several sources, including hereditary, abusive, environmental, and psychological ones. The World Health Organization (WHO) predicted that by 2023, there would be over 280 million depressed people of all ages in the world. Here, females are more likely than males to struggle with depression.

Despite having varying effects in various nations and regions, depression was the second-leading cause of disability worldwide[1]. As an illustration, Afghanistan had the highest rate of serious depression. Due to decreased productivity and higher healthcare expenditures, depression has a huge effect on the world economy[2]. Depression is a significant factor in the nearly 800k suicide deaths that occur each year globally and for every suicide, there are over 20 attempts[3].

In Bangladesh, around 7 million people were thought to have a serious depressive illness [1]. Another study conducted among school-going students in this country reported a prevalence of depression and anxiety [2].

The prevalence rate for anxiety and depression from moderate to severe levels was 26.5% and 18.1%, respectively. Due to missed workdays and decreased production, depression has a substantial influence on a country's economy and productivity [3]. In 2016, the economic cost of depression was estimated by research in the Asian Journal of Psychiatry to be around 4.4% of GDP (gross domestic product). Depression is a significant risk factor for suicide [4]. Suicide rates were 7.3 (95% Confidence Interval, CI 5.6-9.5) per 100,000 per year, with the greatest rate seen in the age group of 60 years and above. In comparison to urban populations, the rate of suicide was found to be 17 times higher (95% CI 5.36-54.64) in rural areas.

In every depression-related suicide, we do not just lose a person. It severely hurts the victim's family and society from a financial point of view. It also interferes with the daily activities of the victim's surrounding people. Moreover, depressed individuals create an unstable work environment, which hinders the progress of an organization. It needs to counsel the depressed people according to their level of depression such that they come back to their regular life. Therefore, it is essential to identify depressed people with the severity level of depression.

[1] https://www.bbc.com/news/health-24818048
[2] https://www.who.int/redirect-pages/mega-menu/emergencies
[3] https://www.who.int/health-topics/suicide

*Corresponding Author Email Address: mnhshisir@gmail.com

Different social networking sites, such as Facebook, Twitter, and others, have become firmly integrated with human life in recent years because of the widespread usage of the internet. Depressed people often use these platforms to express their personal feelings and opinions. Therefore, we focus on Bengali text data gathered from popular social networking sites like Facebook. As a high-inflectional and low-resource language, detecting a Bengali depressive text is a more challenging job. We have identified several pieces of research on depressive texts in the Bengali language. Most of these studies focused on binary classification problems (depression or non-depression) [5]-[10]. Moreover, the authors developed datasets from various emotional points of view like anger, joy, fear, surprise, disgust, sadness, and others for multi-class classification problems [11],[12]. Our following research contributions address the above issues:

Developing a depressive dataset labeling with four classes: non-depressive, mild, moderate, and severe, by verifying a psychology expert.

Experimenting with pre-processing, feature selection, and various machine learning (ML) models, and show the comparative analysis.

Building a better classification system with a transformer-based model for identifying four levels of depression.

We have structured the rest of the paper as follows. In section 2, we have discussed the approaches with merits and demerits of the analogous depressive text detection systems. Section 3 describes our working process, including dataset creation, pre-processing tasks, feature selection, and explanation of ML models. Then, we specify the experimental configuration and discuss the implementation process in section 4. In section 5, we have displayed implementation results and elaborated on comparative performance analysis. Finally, section 6 states a concluding remark and a plan for future research.

2 Related Work

Finding depressive people is one of the primary concerns in society. Researchers are working in this regard by identifying the depression-related text. Unlike English and other high-resource languages, we have detected several small works for identifying Bengali depressive text. We illustrate this research works below:

Khan et al. [5] experimented on small dataset labeling with happy or sad. In the pre-processing phase, the authors replaced contractions with complete forms, removed less significant characters and stop-words, and applied stemming and tokenization. They did not specify feature selection methods to extract appropriate features. They obtained the highest accuracy (98.00%) using the Long Short-Term Memory (LSTM) model for identifying depression-related text. However, the authors presented very fewer discussions about the error analysis of the detection system.

Mohammed et al. [6] worked on a binary classification problem to detect depressive text. At first, the authors discarded numeric values, punctuation, and stop-words and performed stemming. Then they balanced the dataset using a random under-sampling technique. After that, they utilized an Extra Tree (ET) classifier to extract the features and applied the Principal Component Analysis (PCA) method to reduce feature dimension. They got the best accuracy of 92.80% and the F1-score of 93.61% by exploiting the eXtreme Gradient Boost (XGB) algorithm. The authors are concerned with extracting the contextual meaning of the words in a sentence in their detection system.

Mumu et al. [7] analyzed depressive data to classify it into two classes: depressive and non-depressive. After collecting the dataset, they removed emoticons, punctuation, URLs, and stop-words and applied stemming. Finally, they achieved the highest accuracy of 81.49% by utilizing the Logistic Regression (LR) model with TF-IDF (term frequency-inverse document frequency) vectorization. The authors considerably less discuss the anomaly of their detection system.

Uddin et al. [8] successfully identified a depressive text (depressive or non-depressive) with 86.3% accuracy by exploiting the LSTM model with parameter tuning. In the pre-processing stage, they filtered all characters except the alphanumeric characters. The authors showed less attention to feature selection tasks.

Hossen et al. [11] examined depressive text from various points of view, such as emotional aspects like joy, anger, fear, sadness, and others, dimensions of depression including non-depression, mild, moderate, and severe, just identifying the depressive text with depressive or non- depressive, and so on. They cleaned emoticons, words, stop-words, and non-Bengali characters from the dataset. Then they applied tokenization and normalization operations. The authors obtained the best result of 46.64% F1- score and 48.00% accuracy through the LR model with TF-IDF scoring in four classes of depression. The detection system struggled more (25.00% F1-score) identifying the moderate class depressive text. On the contrary, the authors obtained the highest performance of 80.00% accuracy by exploiting LSTM with word-embedding technique in the binary classification of depressive text. Therefore, we still have sufficient scope to enhance the detection performance of the multi-class classification problem.

Tasnim et al. [9] investigated the depressive dataset, labeling it into two classes: depressive and non-depressive. At first, the authors eliminated special characters, punctuation, and stop words. After that, they used tokenization and stemming operations to the entire dataset. Next, they utilized feature extraction techniques, such as count vectorizer, TF-IDF vectorizer, and word embedding, to select feature vectors. Finally, they applied various ML classifiers, where Decision Tree (DT) outperformed the other models with 97.00% accuracy. The authors experiment with the emoji character set. They observed that emojis with text data show better output than only text data. Although the authors successfully identified the depressive text, they did not deal with the intensity level of the depression. Moreover, the authors did not present a comparative analysis of three feature selection techniques in this investigation.

3 Materials and Methods

Identifying depressed people by detecting depressive text is a challenging task. To do this job, at first, we accumulate depression-related text. Afterward, we apply pre-processing tasks for filtering noisy content. Then, we utilize feature selection techniques to extract pertinent features. Lastly, we manipulate ML models to classify four types of Bengali depressive text. Fig. 1 delineates the overall workflow of the multi-class classification system.

3.1 Dataset Description

Nowadays, people use social media to treat mental satisfaction via helpful sharing as a quick method of communication. In this composition, we have standardized the posts and the comments, which are appropriate to depression. Then, we start the data annotation process after finishing the data collection process. We deal with four types of depression data: mild, moderate, severe, and non-depression, in this research.

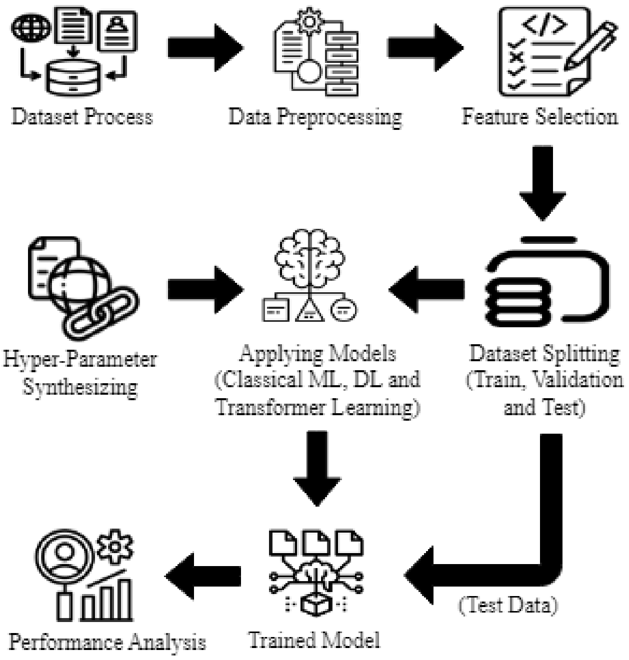

Fig. 1 Overall working process

Mild: It is an initial level of depression characterized by persistent feelings of sadness, hopelessness, and low mood, lasting for at least two weeks. It is also known as dysthymia or persistent depressive disorder. Individuals with mild depression may experience a loss of interest in activities they once enjoyed, have trouble sleeping or sleeping too much, feel fatigued or have low energy, experience changes in appetite or weight, and have difficulty concentrating or making decisions.

Moderate: It is the medium level of depression. People suffering from this depression may have negative thoughts about themselves, their lives, or the future and may experience feelings of worthlessness or guilt.

Severe: It is the extreme level of depression, where people may direct to suicidal thoughts or engage in self-harming behaviors.

Non-depression: The text that does not fall into the above three categories is known as non-depressive text.

We have divided the data collection and annotation process into three phases. In the first phase, we have assigned three annotators who collect data according to the definition (mentioned above) of the four classes of depression. The annotators compile an aggregate of 3,000 data points from various Facebook pages and personal profiles with their endeavored diligence. In the second phase, we distribute these data points by changing between the three annotators. After the second time annotation, we filter the mismatched-label data and get 2,598 data points. In the final stage, we examined these data points with a psychologist to correct mislabeling data. We provide an example of each label of depression in Table 1 and present some statistical information about the dataset, including the number of entries (*quantity*), percentage of each class (*percent*), minimum (*minWords*), maximum (*maxWords*), and average (*avgWords*) number of words in a text, in Table 2.

3.2 Data Pre-processing

In the pre-processing stage, we have removed various unwanted and noisy elements. At first, 118 duplicated instances are discarded. Then, HTML tags, URLs, punctuation marks, special characters, and digits are removed from the dataset. After that, we eliminated all other languages except Bengali. Lastly, we have performed stop-words removal and stemming [13] operations to the entire dataset. Throughout the empirical experiment, we have observed that stop-word removal and stemming operations degrade the overall performance of the classification system. In addition, retaining stop-words and not performing stemming leads to more variety of features [14],[15]. For that reason, these two pre-processing tasks are not considered in the final experiment.

Table 1 Samples of the depression dataset.

Text	**English Translation**	**Label**
কোন জিনিসই অতিরিক্ত হওয়া ভাল নয় দুটি জিনিস ছাড়া। এক: জ্ঞান দুই: ভদ্রতা।	Nothing is decent in excess except two things. One: Knowledge, Two: Gentleness.	Non-depressive
আটকে রাখার মানুষ অনেক! আগলে রাখার মানুষ কই?	There are a lot of people to hold back! Where are the people to keep up?	Mild
নিজের ঘরেই যে মূল্যহীন, পুরো পৃথিবীর কাছে তার মূল্য থাকলেও নিজেকে নিঃস্ব বলেই মনে হয়।	He, who is worthless in his own house, though he has importance to the whole world, feels himself destitute.	Moderate
ইদানিং আমি আমার নিজের রাগকে কন্ট্রোল করতে পারিনা যখন রাগ উঠে তখন আমি আমার ৪ বছরের ছেলেটাকে অনেক মারি, এমন মারা মারি যে মনে হয় সৎ মা ও তাকে এভাবে মারত না, পরে নিজে কান্না করি নিজেরই খারাপ লাগে। ছোট বাচ্চা টাকে গায়ে হাত তুলতে দ্বিধা বোধ করি না, নিজের আত্মবিশ্বাসটা খুবই কম কাউকে সহজে বিশ্বাস করতে পারিনা।	I cannot control my anger when I get angry. I beat my 4-year-old son a lot; I feel like any stepmother would not beat him like that, and then I cry and feel regret for myself. Sometimes I don't hesitate to beat a small child; my self-confidence is very low, and I can't trust anyone easily.	Severe

Table 2 Statistical information of the dataset.

Class	***quantity***	***percent***	***minWords***	***maxwords***	***avgwords***
Non-depressive	949	36.56%	4	115	19
Mild	775	29.85%	4	172	23
Moderate	608	23.42%	4	261	34
Severe	264	10.17%	4	245	40

3.3 Feature Selection

At this stage, we have extracted features from the pre-processed data. Here, we do this task separately for the three types of ML models, classical ML, DL, and transformer-based pre-trained models. For classical ML models, we have utilized character, word, and combinations of character-word N-gram techniques with TF-IDF scoring. Through the experimental observation, we have fixed the value of N as 3 to 5 for the character N-gram and 1 to 2 for the word N-gram. In the case of

DL models, we have observed through empirical analysis that the fastText word embedding approach gives better output than the word2vec and GloVe methods. The main reason is that the fastText employs the character N-gram technique to get sub-word level features, which handle the unknown word vocabulary of the dataset [16],[17]. Here, we have used the *max_len* (the maximum number of tokens in each text) is 132 and the *vector_size* (the size of the feature vector) is 100 to maintain the same input dimension. In the case of transformer-based models, they use their own embedding techniques, where the BERT uses the WordPiece[4] technique and the XLM-R utilizes the Sentence Piece Model (SPM) [18] method to extract token-based sub-word level features. The SPM combines two sub-word segmentation methods: the uni-gram language model [19] and byte-pair-encoding (BPE) [20].

3.4 Model Classifiers

In this experiment, we have chosen nine machine learning models, MNB, SVM, RF, LR, LSTM, BiLSTM, CNN-BiLSTM, BERT, and XLM-R models, which show better performance in several Bengali text classification tasks. The authors of [10],[11],[21], and [22] proposed MNB, LR, SVM, and RF, respectively among the classical ML models for depression or sentiment-related text classification tasks. On the contrary, Khan et al. [5], Tasnim et al. [9], and Mumu et al. [7] achieved better outputs through the LSTM, BiLSTM, and CNN-LSTM-based DL models. However, in recent times, the BERT and XLM-R-based pre-trained models present outstanding performance in the text classification task for low-resource languages [23],[24]. We have explained each model as follows:

Multinomial Naive Bayes (MNB): MNB utilizes the basic principle of the Bayes theorem [25]. Every feature and class variable is computed in the training process through this theorem with prior and conditional probability formulas. Then, the test dataset is subjected to these prior and conditional probabilities. MNB then determines the likelihood for each class of each data point based on the characteristics of the test data and chooses a class with a greater probability. The probability, P(C|F), is measured through the following Bayes theorem:

$$P(C|F) = \frac{P(F|C) * P(C)}{P(F)}$$

where C represents a class variable, and F denotes the distinctive feature.

Support Vector Machine (SVM): By maximizing the distance across the margins of the two types of support vectors (like positive and negative), SVM creates a hyper-plane space. This marginal distance leads to the production of a more broadly applicable model. SVM also addresses errors by applying a regularization technique to incorrectly categorized data points that are based around the soft margin hyper-plane [26].

Random Forest (RF): The RF comes by combining several decision trees that are computed individually [27]. Here, the "Gini Impurity" formula is used to obtain the information gained for each tree as follows:

$$Gini = 1 - \sum_{i=1}^{N} (P_i)^2$$

where N is the total number of possible classes, and P_i is the probability of i^{th} class.

Logistic Regression (LR): The operational procedure of this model consists of the sigmoid and cost functions, calculated as Eqs. 1 and 2, respectively.

$$h_\theta(x) = \frac{1}{1 + e^{-\beta^T x}} \tag{1}$$

$$C(\theta) = \frac{1}{j}\sum_{i=1}^{j} c(h_\theta(x_i), y_i) \tag{2}$$

$$c(h_\theta(x), y) = \begin{cases} -\log(1 - h_\theta(x)), & if\ y = 0 \\ -\log(h_\theta(x)), & if\ y = 1 \end{cases}$$

where, β^T denotes the transpose of regression coefficient matrix (β), j indicates the total training observations, $h_\theta(x_i)$ represents the hypothesis function of the i^{th} training sample, and y_i gives the true output of the i^{th} training observation.

Long Short-Term Memory (LSTM): This paradigm can detain long-term dependencies in sequential data such as time series, text, and speech [28]. They manage the flow of information using a memory cell and gates, enabling them to keep or get rid of information as needed. The input gate, forget gate and output gate are all connected to the memory cell in different ways. The input gate selects the information that needs to be stored in the memory cell. The memory cell's information is selected for discard by the forget gate. As a last step, the output gate unit selects the data that will be sent to the subsequent LSTM cell state.

Bidirectional LSTM (BiLSTM): A Bidirectional LSTM [29] is a sequential data processing model that has two LSTM layers, one of which processes the input sequence forward and the other of which processes it backward. The final forecast is created by merging the results from the two directions, with each layer maintaining its own hidden states. Additionally, BiLSTMs effectively expand the network's pool of knowledge, giving the algorithm better context.

Convolutional Neural Network with BiLSTM (CNN-BiLSTM): A Convolutional Neural Network with Bidirectional Long Short-Term Memory (CNN-BiLSTM) is a hybrid model that combines the unique features of both convolutional neural networks (CNNs) and bidirectional LSTM (BiLSTM) networks. The CNN-BiLSTM architecture consists of three main features: The CNN consists of one or more convolutional layers, followed by pooling layers and activation functions. Convolutional layers apply filters to the input data, rotate the spatial dimensions, and create feature maps that illustrate various facets of the input. The network may gather context data for the past and the future thanks to its bidirectional processing. The final sequence representation is created by concatenating the outputs of the forward and backward LSTMs. After the CNN-BiLSTM levels, fully connected layers apply non-linear transformation changes to all the neurons from the preceding layers and connect them to the output layer, where they produce the final predictions or classifications [30].

Bidirectional Encoder Representations from Transformers (BERT): BERT [31] can be fine-tuned for a variety of downstream natural language processing (NLP) tasks, including question-answering, named entity recognition, and sentiment

[4] https://huggingface.co/course/chapter6/6?fw=pt

analysis. It is pre-trained on substantial volumes of unlabeled text data. During the pre-trained, BERT uses the Next Sentence Prediction (NSP) and Mask Language Model (MLM) unsupervised tasks to learn the context of the specified languages. BERT takes into consideration both the left and right contexts of each word in a phrase simultaneously, unlike traditional models that analyze text sequentially (left-to-right or right-to-left).

Cross-lingual Language Model with Robustly Optimized BERT (XLM-RoBERT): XLM-RoBERTa [24] is a sophisticated language model with numerous layers and features. Each encoder layer consists of sub-layers, including multi-head self-attention and position-wise feed-forward networks, which allow the model to capture the contextual relationships between words in the input sequence. A series of tokens encoding the text is commonly used as the input to XLM-RoBERTa. Using SPM [18], these tokens are first transformed into numerical embeddings. The embeddings accurately depict each token's semantic meaning within the context of the phrase. It utilizes cross-lingual pre-training to develop its ability to comprehend and produce text in several languages. The model is trained on a sizable corpus of monolingual data from many languages during pre-training. This enables it to acquire representations independent of a language and to identify the linguistic traits that are common to all languages. After pre-training, the model is more trained on certain downstream tasks, such as the classification of texts, named entity identification, or machine translation. The model's knowledge is adjusted during this step of fine-tuning, which also aids in enhancing the model's performance on the targeted language-related tasks.

4 Experimental Set-up and Implementation

The deep learning and the transformer-based models need a highly configured processing environment. For this purpose, we have used the Google Colab cloud environment, which provides the Jupyter Notebook Python programming language editor and felicitates NVIDIA Graphics Processing Units (GPU) with many built-in Python modules and packages [32]. In this case, the GPU is a Tesla T4 with 15GB of RAM. The experimental configuration of this work is discussed from three aspects, for classical ML, DL, and transformer-based models. Regarding classical ML models, at first, the dataset is split into two parts: train and test with a ratio of 90% and 10%, respectively. After that, sklearn python packages are used with default hyper-parameter values to execute each model. During the execution, a 10-fold cross-validation approach is incorporated to get a more reliable classification system [33]. In the case of DL models, we have split the dataset into train, validation, and test with a ratio of 70%, 20%, and 10%, respectively. We have utilized tensorflow Python packages to implement DL models. We use two LSTM layers, one BiLSTM layer, and one CNN and one BiLSTM layer for implementing LSMT, BiLSTM, and CNN-BiLSTM models, respectively. In every model, we add two hidden (dense) layers and one output layer. In addition, a dropout function is utilized between every two layers to handle the overfitting problem [34]. Throughout the empirical experiment, the overall parameter values of DL models are settled which are figured out in Table 3. On the other hand, we utilize ktrain python packages to implement transformer-based models. We use the same dataset splitting ratio as DL models. We have tried various values of *max_len*, *learning_rate*, and *batch_size*. We obtained the best-configured values for the *learning_rate* of 4e-05 and the *batch_size* of 12 for the bert-base-multilingual-uncased and the xlm-roberta-base models. However, *max_len* is set as 120 for the BERT-based model and 124 for the XLM-R-based model. In this work, we run each model up to 10 epochs.

Table 3 Hyper-parameter values during the implementation of deep learning models.

Parameter	**Data Type**	**Description**	**Value**
LSTM units	Integer	The amount of LSTM output units.	140
BiLSTM units	Integer	The amount of BiLSTM output units (for the BiLSTM model) in each direction.	140
Filters	Integer	The number of convolution filters.	192
Kernel size	Integer	The length of the convolution window.	3
Pooling types	String	Two types: max pooling and average pooling, are used to reduce the dimension of the feature map.	max pooling’
BiLSTM units	Integer	The amount of BiLSTM output units (for the CNN-BiLSTM model) in each direction.	128
Hidden units	Integer	The number of neurons of each dense layer.	128
Activation function	String	The function defines the output value of each neuron.	’relu’
Kernel initializer	String	A procedure to assign a set of small random values of a neural network at the very early stage.	’glorot_uniform’
Dropout rate	Float	Fraction of the number of neurons to drop.	0.2
Learning rate (Adam)	Float	It controls how fast a loss function moves toward a point where the curves meet.	0.0001
Batch size	Integer	The number of samples participating in each iteration during training of the model.	12
Epochs	Integer	The number of times all training data is utilized to train the model.	30

5 Result and Discussion

For measuring the performance of each ML model, we have taken four evaluation metrics: precision, recall, F1-score, and accuracy. We use the weighted average value for each metric. We have illustrated this section from four aspects: results of classical ML models, results of DL models, results of transformer-based models, and overall result analysis. First, we discuss the comparative analysis of classical ML models. Subsequently, we compare the outputs of the DL models. We then analyze the results of the transformer-based pre-trained models. Lastly, we present the overall analysis of the ML models.

5.1 Result of Classical ML Models

The outputs of four classical ML models: SVM, RF, LR, and MNB are articulated in Table 4. Here, the SVM with character N-grams features obtains the highest performance in precision, recall, F1-score, and accuracy of 54.83%, 54.45%, 52.13%, and 54.45%, respectively. Character N-grams give better accuracy than the word N-grams and the combined N-grams to the other three models: 52.47% for MNB, 51.86% for RF, and 52.83% for LR. The main reason behind this performance is that character N-grams technique produces many sub-words level features, which play an important role for classifying Bengali depressive text.

Table 4 Performance evaluation of the classical ML models

Approach	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**Accuracy (%)**
Character N-grams + SVM	**54.83**	**54.45**	**52.13**	**54.45**
Word N-grams + SVM	50.95	50.82	47.16	50.82
Combined N-grams + SVM	53.24	53.32	51.65	53.32
Character N-grams + MNB	53.12	52.47	50.66	52.47
Word N-grams + MNB	48.83	48.23	47.69	48.23
Combined N-grams + MNB	52.89	51.70	51.17	51.70
Character N-grams + RF	51.22	51.86	47.54	51.86
Word N-grams + RF	48.69	46.98	40.76	46.98
Combined N-grams + RF	50.68	50.25	45.94	50.25
Character N-grams + LR	53.20	52.83	49.30	52.83
Word N-grams +LR	49.06	49.36	43.46	49.36
Combined N-grams + LR	52.62	52.79	49.40	52.79

5.2 Result of DL Models

We notify the implementation results of three DL models: LSTM, BiLSTM, and CNN-BiLSTM in Table 5. The BiLSTM model obtains the top score in the four-evaluation metrics: 58.96% of precision, 58.87% of recall, 56.50% of F1-score, and 58.87% of accuracy. This DL model gets contextual understanding from a sentence with forward and backward directions, which are the principal reasons for the better performance, compared to the other two DL models. However, the CNN-BiLSTM model shows a similar output to BiLSTM with a slight decrease in value. Since the LSTM model works only forward direction, the F1-score and accuracy give a lower value of 52.75% and 56.05%, respectively.

Table 5 Performance evaluation of the DL models

Approach	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**Accuracy (%)**
LSTM	50.77	56.05	52.75	56.05
BiLSTM	**58.96**	**58.87**	**56.50**	**58.87**
CNN-BiLSTM	58.52	58.06	55.86	58.06

Table 6 Performance evaluation of the transformer-based models

Approach	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**Accuracy (%)**
bert-base-multilingual-uncased	59.45	56.85	56.87	56.85
xlm-roberta-base	**61.63**	**60.89**	**61.11**	**60.89**
LR + TF-IDF [15]			46.64	48.00

5.3 Result of Transformer-based Models

We figure out the performance of two transformer-based models: BERT and XLM-R, in Table 6. The XLM-R-based approach achieves better results compared to the BERT-based method and the earlier work [11] in terms of precision (61.63%), recall (60.89%), F1-score (61.11%), and accuracy (60.89%). The XLM-R is pre-trained over a large dataset (two terabytes) from one hundred languages, including low-resource languages like Bengali. Furthermore, this pre-trained model utilizes the SPM technique with a larger vocabulary (250k) to get sub-word level feature vectors [24]. For these reasons, the XLM-R-based approach shows much better performance for detecting depressive text.

5.4 Overall Result Analysis

By observing Table 4-Table 6 we see that the XLM-R-based approach outperforms the other ML models in this research. This model achieves the best score in all four-evaluation metrics such as precision, recall, F1-score, and accuracy. For that reason, we only focus on this model in the subsequent analysis.

Fig. 2 displays the training and validation accuracy curve for the XLM-R-based approach. The training curve gradually increases from 0.35 to 0.90 until the tenth epoch. On the contrary, the validation curve shows ups and downs with an upward trend, and the final accuracy value reaches 0.56 at the tenth epoch.

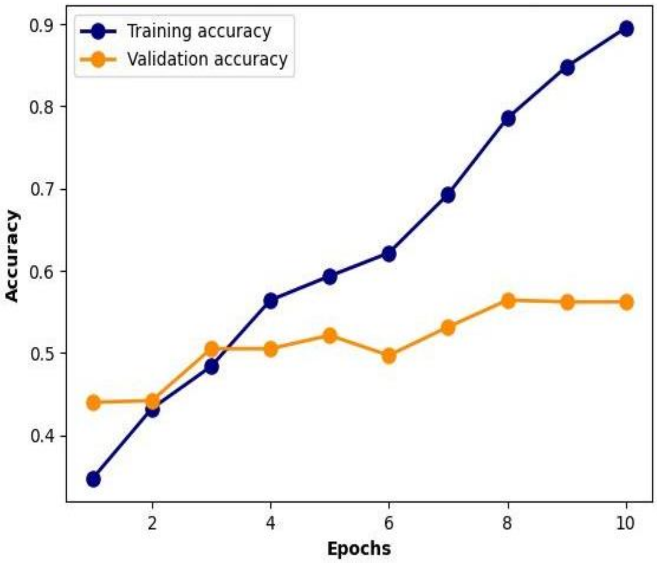

Fig. 2 Training and validation accuracy curve

We also visualize the training and validation loss in Fig. 3. The training loss moderately reduces from the first (loss value is 1.32) to the tenth (loss value is 0.31) epoch. However, the validation loss presents a different scenario. In the third epoch, we count the lowest loss as 1.07. After that, the curve slowly rises until the last epoch, and the final loss value reaches 1.47.

We now observe the class-wise performance of the depressive text dataset (Table 7).

Table 7 Class-wise performance measure of the XLM- R-based approach

Class	**Precision (%)**	**Recall (%)**	**F1-score (%)**
Non-depressive	77.38	73.03	75.14
Mild	49.43	57.33	53.09
Moderate	55.77	50.00	52.73
Severe	56.00	53.85	54.90

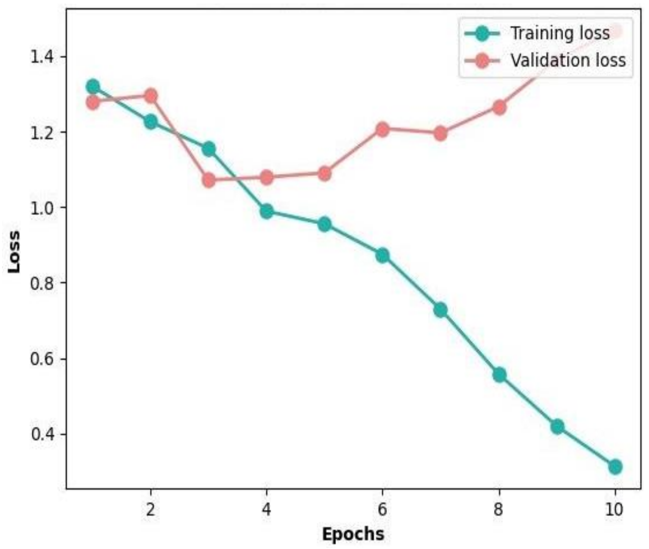

Fig. 3 Training and validation loss curve

Since the Non- depressive class comprises the most data (36.56%), it performs much better in the precision of 77.38%, recall of 73.03%, and F1-score of 75.14% than the other three classes. The Mild, Moderate, and Severe classes give nearly similar F1-scores of 53.09%, 52.73%, and 54.90%, respectively.

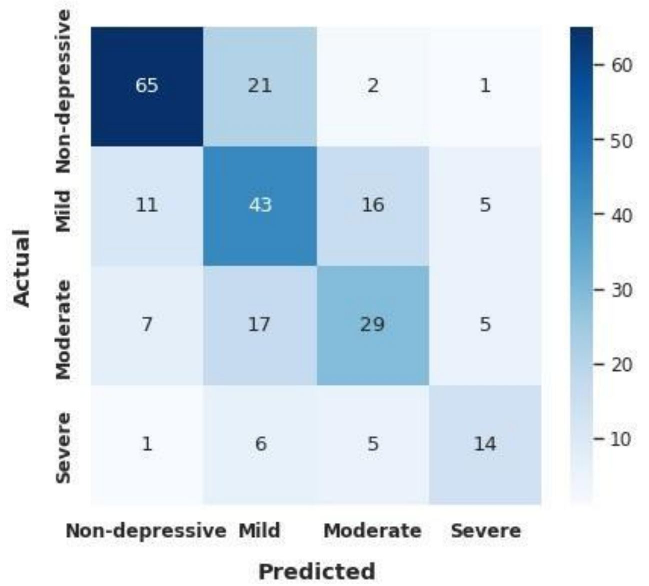

Fig. 4 Confusion matrix of the XLM-R-based detection system

The highest precision and recall value among these three classes is 56.00% for the severe class and 57.33% for the Mild class. Now, we visualize biases and misclassification issues of every class in Fig. 4. The Non- depressive is mostly biased by the Mild class, and vice-versa. The Mild and Moderate categories are also highly influenced by each other. And the Severe class overlaps with the Mild and Moderate groups. In a nutshell, the three classes: Mild, Moderate, and Severe, have more inclinations to overlap with one another. Thus, we get lower performance scores for these three classes (Table 7).

6 Conclusion

Our proposed XLM-R-based approach outperforms the existing works for the multi-class classification of the Bengali depressive text. We have tried various feature selection techniques and ML models in this research. Finally, we observe that the XLM-R-based approach successfully detects a depressive text with a 61.11% F1-score and 60.89% accuracy. However, there is still enough space to enhance the detection system. Since we experiment with a small dataset, we will enlarge our dataset in the subsequent research. In addition, we will handle the overlapping issue of the three depression classes: Mild, Moderate, and Severe. Lastly, we will concentrate on pre-processing and feature selection levels and apply other cutting-edge ML models to improve the detection performance of the depressive text.

Acknowledgment

The authors are thankful to Mrs. Farhana Yeasmin Satu, a Psychologist in Bangladesh, who provides expert opinions in the entire process of the dataset development. Mrs. Satu completed her Bachelor of Science and Master of Science degrees from the University of Chittagong, Bangladesh. She achieved a certificate in Mental Health, Depression, Anxiety, and Autism from the United States of America (USA).

References

[1] Arusha, A.R. and Biswas, R.K., 2020. Prevalence of stress, anxiety and depression due to examination in Bangladeshi youths: A pilot study. Children and youth services review, 116, p.105254.

[2] Islam, M.S., Rahman, M.E., Moonajilin, M.S. and van Os, J., 2021. Prevalence of depression, anxiety and associated factors among school going adolescents in Bangladesh: Findings from a cross-sectional study. Plos one, 16(4), p.e0247898..

[3] Ogbo, F.A., Mathsyaraja, S., Koti, R.K., Perz, J. and Page, A., 2018. The burden of depressive disorders in South Asia, 1990–2016: findings from the global burden of disease study. BMC psychiatry, 18(1), pp.1-11.

[4] Mashreky, S.R., Rahman, F. and Rahman, A., 2013. Suicide kills more than 10,000 people every year in Bangladesh. Archives of Suicide Research, 17(4), pp.387-396.

[5] Rafidul Hasan Khan, M., Afroz, U.S., Masum, A.K.M., Abujar, S. and Hossain, S.A., 2021. A deep learning approach to detect depression from Bengali text. In Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2020, Volume 2 (pp. 777-785). Springer Singapore.

[6] Mohammed, M.B., Abir, A.S.M., Salsabil, L., Shahriar, M. and Fahmin, A., 2021, December. Depression Analysis from Social Media Data in Bangla Language: An Ensemble Approach. In 2021 Emerging Technology in Computing, Communication and Electronics (ETCCE) (pp. 1-6). IEEE.

[7] Mumu, T.F., Munni, I.J. and Das, A.K., 2021. Depressed people detection from bangla social media status using lstm and cnn approach. Journal of Engineering Advancements, 2(01), pp.41-47.

[8] Uddin, A.H., Bapery, D. and Arif, A.S.M., 2019, July. Depression analysis from social media data in Bangla language using long short term memory (LSTM) recurrent neural network technique. In 2019 International Conference on Computer, Communication,

Chemical, Materials and Electronic Engineering (IC4ME2) (pp. 1-4). IEEE.

[9] Tasnim, F., Habiba, S.U., Nafisa, N. and Ahmed, A., 2022. Depressive Bangla text detection from social media post using different data mining techniques. In Computational Intelligence in Machine Learning: Select Proceedings of ICCIML 2021 (pp. 237-247). Singapore: Springer Nature Singapore.

[10] Khan, M.R.H., Afroz, U.S., Masum, A.K.M., Abujar, S. and Hossain, S.A., 2020, July. Sentiment analysis from bengali depression dataset using machine learning. In 2020 11th international conference on computing, communication and networking technologies (ICCCNT) (pp. 1-5). IEEE.

[11] Hossen, I., Islam, T., Rashed, M.G. and Das, D., 2022, October. Early Suicide Prevention: Depression Level Prediction Using Machine Learning and Deep Learning Techniques for Bangladeshi Facebook Users. In Proceedings of International Conference on Fourth Industrial Revolution and Beyond 2021 (pp. 735-747). Singapore: Springer Nature Singapore.

[12] Das, A., Sharif, O., Hoque, M.M. and Sarker, I.H., 2021. Emotion classification in a resource constrained language using transformer-based approach. arXiv preprint arXiv:2104.08613.

[13] Seddiqui, M.H., Maruf, A.A.M. and Chy, A.N., 2016. Recursive suffix stripping to augment bangla stemmer. In International Conference Advanced Information and Communication Technology (ICAICT).

[14] Ahmed, M.T., Rahman, M., Nur, S., Islam, A. and Das, D., 2021, February. Deployment of machine learning and deep learning algorithms in detecting cyberbullying in bangla and romanized bangla text: A comparative study. In 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) (pp. 1-10). IEEE.

[15] Kumar, R., Lahiri, B. and Ojha, A.K., 2021. Aggressive and offensive language identification in hindi, bangla, and english: A comparative study. SN Computer Science, 2(1), p.26.

[16] Joulin, A., Grave, E., Bojanowski, P. and Mikolov, T., 2016. Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759.

[17] Grave, E., Bojanowski, P., Gupta, P., Joulin, A. and Mikolov, T., 2018. Learning word vectors for 157 languages. arXiv preprint arXiv:1802.06893.

[18] Kudo, T. and Richardson, J., 2018. Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. arXiv preprint arXiv:1808.06226.

[19] Kudo, T., 2018. Subword regularization: Improving neural network translation models with multiple subword candidates. arXiv preprint arXiv:1804.10959.

[20] Sennrich, R., Haddow, B. and Birch, A., 2015. Neural machine translation of rare words with subword units. arXiv preprint arXiv:1508.07909.

[21] N. R. Bhowmik, M. Arifuzzaman, M. R. H. Mondal, and M. Islam, "Bangla text sentiment analysis using supervised machine learning with extended lexicon dictionary," Natural Language Processing Research, vol. 1, no. 3-4, pp. 34–45, 2021.

[22] Tabassum, N. and Khan, M.I., 2019, February. Design an empirical framework for sentiment analysis from Bangla text using machine learning. In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE) (pp. 1-5). IEEE.

[23] Jahan, M.S., Haque, M., Arhab, N. and Oussalah, M., 2022, June. BanglaHateBERT: BERT for Abusive Language Detection in Bengali. In Proceedings of the Second International Workshop on Resources and Techniques for User Information in Abusive Language Analysis (pp. 8-15).

[24] Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L. and Stoyanov, V., 2019. Unsupervised cross-lingual representation learning at scale. arXiv preprint arXiv:1911.02116.

[25] Xu, S., Li, Y. and Wang, Z., 2017. Bayesian multinomial Naïve Bayes classifier to text classification. In Advanced Multimedia and Ubiquitous Engineering: MUE/FutureTech 2017 11 (pp. 347-352). Springer Singapore.

[26] Cortes, C. and Vapnik, V., 1995. Support-vector networks. Machine learning, 20, pp.273-297.

[27] Ali, J., Khan, R., Ahmad, N. and Maqsood, I., 2012. Random forests and decision trees. International Journal of Computer Science Issues (IJCSI), 9(5), p.272.

[28] Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. Neural computation, 9(8), pp.1735-1780.

[29] Schuster, M. and Paliwal, K.K., 1997. Bidirectional recurrent neural networks. IEEE transactions on Signal Processing, 45(11), pp.2673-2681.

[30] Karim, M.R., Chakravarthi, B.R., McCrae, J.P. and Cochez, M., 2020, October. Classification benchmarks for under-resourced bengali language based on multichannel convolutional-lstm network. In 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA) (pp. 390-399). IEEE.

[31] Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[32] Carneiro, T., Da Nóbrega, R.V.M., Nepomuceno, T., Bian, G.B., De Albuquerque, V.H.C. and Reboucas Filho, P.P., 2018. Performance analysis of google colaboratory as a tool for accelerating deep learning applications. IEEE Access, 6, pp.61677-61685.

[33] Wong, T.T. and Yeh, P.Y., 2019. Reliable accuracy estimates from k-fold cross validation. IEEE Transactions on Knowledge and Data Engineering, 32(8), pp.1586-1594.

[34] Baldi, P. and Sadowski, P.J., 2013. Understanding dropout. Advances in neural information processing systems, 26.

This page is left intentionally blank

Journal of Engineering Advancements (JEA)

DOI: https://doi.org/10.38032/jea

Indexed by:

Google Scholar Crossref

Scilit

Volume 04 Issue 02

DOI: https://doi.org/10.38032/jea.2023.02

Published by: SciEn Publishing Group

Website: www.scienpg.com